A Systems Analysis of the Spam Problem

by

Gabriel R. Weinberg
Bachelor of Science, Physics
Massachusetts Institute of Technology

Submitted to the Engineering Systems Division
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Technology and Policy
at the
Massachusetts Institute of Technology

June 2005

©2005 Massachusetts Institute of Technology. All rights reserved.

Signature of Author
April 20, 2005

Certified by
Sharon Eisner Gillett
Principal Research Associate, Center for Technology, Policy and Industrial Development
Thesis Supervisor

Certified by
David Clark
Senior Research Scientist, Computer Science and Artificial Intelligence Laboratory
Thesis Supervisor

Accepted by
Dava J. Newman
Professor of Aeronautics and Astronautics and Engineering Systems
Director, Technology and Policy Program

ARCHIVES
Abstract

This thesis considers the problem of the large amount of unwanted email that is being sent and received, which lowers the aggregate value of email as a communication medium from what it would otherwise be. This problem is commonly known as the "spam problem."

Solutions to the spam problem involve curbing the adverse affects of existing technology as well as steering technology development in a socially beneficial direction. Unlike some other technology and policy problems, the reasons for the existence of this problem are well known and the desired effects of ideal solutions can be readily articulated. However, attempted solutions to date have not made much progress at solving the problem. I posit that this failure stems from the fact the spam problem is really a complex system, and that solutions to date have not been designed to interact with this system in a useful manner.

I show that the spam problem is a complex system, and should be dealt with by developing strategies to holistically interact with it. Such strategies must embrace both technical and legal realities simultaneously in order to be successful. They must also avoid causing negative side effects that negate their purpose.

First, I build a model of the system surrounding the spam problem in the form of a Causal Loop Diagram. This diagram shows the causal interactions between the various technical, legal, social, and economic forces that are present in the spam system. Using this diagram, I then identify a number of places that solutions could interact with this system. These places comprise a set of possible levers that could be pulled to alleviate the spam problem.

This set of levers is then used to make sense of the attempted and suggested solutions to date. Various solutions are grouped by how they interact with the system. These solution categories are then presented in detail by showing, diagrammatically, how they positively and negatively affect the spam system through their interactions with it. In so doing, I attempt to argue persuasively that much of the current energy expended toward the spam problem is largely unnecessary, and in some cases, counterproductive.

I additionally argue that because of the current reality of the spam problem, i.e. particular facts, we are already in a decent position to largely solve this problem by just redirecting current efforts toward more appropriate activity. Such appropriate activity is suggested, which includes steps to increase the identifiability of email in order to enable more successful litigation. Finally, an optimistic conclusion is reached that there are no fundamental reasons why the spam problem can not be dealt with in such a manner to ensure the continued usefulness of email as a communication medium.
Acknowledgements

I would like to acknowledge the following people and organizations for their roles in helping me produce this thesis.

- Sharon Eisner Gillett
- David Clark
- Charlie Fine
- Lauren McCann
- Daniel Cunningham
- Joseph Sussman
- David Zipkin
- Chintan Vaishnav
- Simson Garfinkel
- Lynn and Winkler Weinberg
- Rita Adom, Becky Shepardson and everyone involved with the Communications Future Program at MIT.
- Sydney Miller, Dava Newman, and everyone involved with the Technology and Policy Program at MIT.
Table of Contents

ABSTRACT .......................................................................................................................... 2
ACKNOWLEDGEMENTS ...................................................................................................... 3
TABLE OF CONTENTS ........................................................................................................ 4
TABLE OF TABLES .................................................................................................................. 6
TABLE OF FIGURES ........................................................................................................... 7

1 INTRODUCTION ....................................................................................................................... 8
  1.1 THESIS STRUCTURE ............................................................................................................ 8
  1.2 READING THIS THESIS ...................................................................................................... 9
  1.3 WHY THE DEFINITION OF SPAM IS IMPORTANT .............................................................. 9
  1.4 DIFFICULTIES CREATING AN IDEAL DEFINITION OF SPAM ........................................ 10
  1.5 ERRORS MADE WHEN APPLYING REALISTIC DEFINITIONS OF SPAM ............................ 11
  1.6 TRADING PRECISENESS FOR ACCURACY WHEN CREATING REALISTIC DEFINITIONS OF SPAM .................................................................................................................. 12
  1.7 USING FRAUDULENT ASPECTS OF MESSAGES WHEN CREATING PRECISE RULES .......... 12
  1.8 CONSTRUCTING A USEFUL DEFINITION OF SPAM .......................................................... 14
  1.9 WHAT IS SPAM? ................................................................................................................ 16
  1.10 THE SPAM RATIO .......................................................................................................... 17

2 SENDING EMAIL .................................................................................................................... 18
  2.1 BASIC EMAIL PROTOCOLS ............................................................................................. 18
  2.2 SMTP BASICS ................................................................................................................ 20
  2.3 EMAIL FORMAT .............................................................................................................. 22
  2.4 SMTP EXPLOITATION ..................................................................................................... 23
  2.5 IP ADDRESS BASICS ..................................................................................................... 25
  2.6 IP ADDRESS EXPLOITATION ......................................................................................... 26
  2.7 EMAIL EXPLOITATION SUMMARIZED ......................................................................... 28

3 REVIEW OF SPAM SOLUTION ACTIVITY .......................................................................... 29
  3.1 OVERVIEW OF MAJOR SPAM SOLUTION ACTIVITY .................................................... 29
  3.2 CONTENT FILTERING .................................................................................................... 31
  3.3 ACCESS FILTERING ...................................................................................................... 33
  3.4 IDENTIFIABILITY ............................................................................................................ 37
  3.5 REGULATION ................................................................................................................ 39
  3.6 INTERNATIONAL COLLABORATION ............................................................................ 40
  3.7 ENFORCEMENT CAPACITY ......................................................................................... 41
  3.8 SPAM SOLUTION ACTIVITY SUMMARIZED .................................................................... 42

4 BASE SPAM SYSTEM ............................................................................................................ 44
  4.1 CAUSAL LOOP METHODOLOGY .................................................................................... 45
  4.2 SPAM RATIO ................................................................................................................ 46
  4.3 SPAM SENT .................................................................................................................. 48
  4.4 SOCIAL REACTION ...................................................................................................... 51
  4.5 REGULATION ............................................................................................................... 52
  4.6 BASE SPAM SYSTEM .................................................................................................. 54
  4.7 BASE SPAM SYSTEM LEVERS .................................................................................. 56

5 SPAM SOLUTION INTERACTIONS WITH THE BASE SPAM SYSTEM .................................... 58
  5.1 SPAM SOLUTION ACTIVITY REVISITED ..................................................................... 58
  5.2 CONTENT FILTERING .................................................................................................. 59
  5.3 ACCESS FILTERING .................................................................................................... 62
  5.4 IDENTIFIABILITY .......................................................................................................... 66
  5.5 REGULATION ................................................................................................................ 68
  5.6 INTERNATIONAL COLLABORATION ........................................................................... 70
  5.7 ENFORCEMENT CAPACITY ....................................................................................... 71
  5.8 SPAM SOLUTION INTERACTIONS SUMMARIZED ....................................................... 72

6 FORMING EFFECTIVE SPAM SOLUTION STRATEGIES .......................................................... 73
  6.1 KEY SPAM SOLUTION ISSUES ..................................................................................... 73
  6.2 WHAT HAS NOT BEEN TRIED? ..................................................................................... 74
## Table of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Email Properties and Their Relation to Spam</td>
<td>14</td>
</tr>
<tr>
<td>Table 2</td>
<td>Sub-Categories of Email That is Not Fraudulent and Has Working Opt-out</td>
<td>15</td>
</tr>
<tr>
<td>Table 3</td>
<td>Email Exploitation Techniques</td>
<td>28</td>
</tr>
<tr>
<td>Table 4</td>
<td>Major Spam Solution Activity</td>
<td>29</td>
</tr>
<tr>
<td>Table 5</td>
<td>Major Spam Solution Types Revisited</td>
<td>58</td>
</tr>
<tr>
<td>Table 6</td>
<td>Base Spam System Levers and Spam Solution Categories</td>
<td>74</td>
</tr>
<tr>
<td>Table 7</td>
<td>Spam Solution Categories and Negative Side Effects</td>
<td>77</td>
</tr>
</tbody>
</table>
# Table of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIGURE 1</td>
<td>EMAIL PROTOCOLS IN USE</td>
<td>19</td>
</tr>
<tr>
<td>FIGURE 2</td>
<td>SMTP IN USE</td>
<td>20</td>
</tr>
<tr>
<td>FIGURE 3</td>
<td>REVERSE DNS TEST</td>
<td>36</td>
</tr>
<tr>
<td>FIGURE 4</td>
<td>IP ADDRESS AUTHENTICATION</td>
<td>38</td>
</tr>
<tr>
<td>FIGURE 5</td>
<td>CONTENT SIGNING</td>
<td>38</td>
</tr>
<tr>
<td>FIGURE 6</td>
<td>BASE SPAM SYSTEM</td>
<td>44</td>
</tr>
<tr>
<td>FIGURE 7</td>
<td>SPAM RATIO</td>
<td>46</td>
</tr>
<tr>
<td>FIGURE 8</td>
<td>ABILITY TO SPAM</td>
<td>48</td>
</tr>
<tr>
<td>FIGURE 9</td>
<td>OTHER PROFIT LOOPS</td>
<td>50</td>
</tr>
<tr>
<td>FIGURE 10</td>
<td>SOCIAL REACTION</td>
<td>51</td>
</tr>
<tr>
<td>FIGURE 11</td>
<td>REGULATIVE REACTION</td>
<td>52</td>
</tr>
<tr>
<td>FIGURE 12</td>
<td>REGULATIVE RESTRICTION</td>
<td>54</td>
</tr>
<tr>
<td>FIGURE 13</td>
<td>BASE SPAM SYSTEM</td>
<td>55</td>
</tr>
<tr>
<td>FIGURE 14</td>
<td>BASE SPAM SYSTEM LEVERS</td>
<td>56</td>
</tr>
<tr>
<td>FIGURE 15</td>
<td>CONTENT FILTERING EFFECTS</td>
<td>59</td>
</tr>
<tr>
<td>FIGURE 16</td>
<td>CONTENT FILTERING SIDE EFFECTS</td>
<td>60</td>
</tr>
<tr>
<td>FIGURE 17</td>
<td>ACCESS FILTERING EFFECTS</td>
<td>63</td>
</tr>
<tr>
<td>FIGURE 18</td>
<td>ACCESS FILTERING SIDE EFFECTS</td>
<td>64</td>
</tr>
<tr>
<td>FIGURE 19</td>
<td>IDENTIFIABILITY EFFECTS</td>
<td>66</td>
</tr>
<tr>
<td>FIGURE 20</td>
<td>IDENTIFIABILITY SIDE EFFECTS</td>
<td>67</td>
</tr>
<tr>
<td>FIGURE 21</td>
<td>REGULATION EFFECTS</td>
<td>69</td>
</tr>
<tr>
<td>FIGURE 22</td>
<td>INTERNATIONAL COLLABORATION EFFECTS</td>
<td>70</td>
</tr>
<tr>
<td>FIGURE 23</td>
<td>ENFORCEMENT CAPACITY EFFECTS</td>
<td>71</td>
</tr>
<tr>
<td>FIGURE 24</td>
<td>BASE SPAM SYSTEM LEVERS</td>
<td>74</td>
</tr>
<tr>
<td>FIGURE 25</td>
<td>EDUCATION EFFECTS</td>
<td>75</td>
</tr>
<tr>
<td>FIGURE 26</td>
<td>PUBLICITY EFFECTS</td>
<td>76</td>
</tr>
<tr>
<td>FIGURE 27</td>
<td>EXAMPLE STRATEGY</td>
<td>81</td>
</tr>
</tbody>
</table>
1 Introduction

This thesis considers the problem of the large amount of unwanted email that is being sent and received, which lowers the aggregate value of email as a communication medium from what it would otherwise be. This problem is commonly known as the "spam problem." The primary argument of this thesis is that the spam problem is a complex system (termed the Base Spam System), and should be dealt with as such.

The central contribution of this thesis is a model of the Base Spam System. This model takes a diagrammatical form that attempts to capture the various causes and effects present. The model is structured in such a way that it can be expanded by future work as the nature of the spam problem continues to change.

To date, attempted and proposed solution strategies to the spam problem have been less than ideal, although numerous and varied. An argument will be presented that this failure stems from the fact these solution strategies have not been designed to interact with the Base Spam System in a holistic manner.

Consideration is also given to how a more holistic approach to the spam problem might be structured. Such a holistic approach is shown to involve significant coordination between various stakeholder groups involved in the Base Spam System. While this kind of coordination is relatively difficult to achieve, an optimistic conclusion is reached that there is no fundamental reasons why it is not possible.

1.1 Thesis Structure

This thesis is divided into six chapters. This first chapter proceeds by examining the difficulties in defining spam. A definition is then chosen for the purposes of the thesis, centering on the delineation between fraudulent and not fraudulent email sent in good faith.

The second chapter begins with a technical introduction about the process of sending email. Then it explains how this process is exploited by spam senders. General notions on such exploitation are explored.

The third chapter is a review of spam solution activity to date. It is not intended to be an evaluation of solutions, but instead a compendium of the types of attempted and proposed solutions. The third chapter explains how each type basically works so that each type can be considered in more detail in relation to the Base Spam System in the fifth chapter.

The fourth chapter introduces the model of the Base Spam System. The model is a causal loop diagram, which is a type of diagram used in the discipline of system dynamics. The reasons for choosing this methodology are presented as well as an introduction to causal loop diagrams in general. Then the Base Spam System model is built up piece by piece, with

1 SPAM® is also a registered trademark of Hormel Foods, associated with its meat product. In this paper, the word "spam" in all of its capitalizations is used exclusively to refer to Internet spam and not to the meat product.
consideration given to each piece individually. The end of the fourth chapter presents the model in its entirety, and also highlights the levers that can be pulled by solutions in attempt to alleviate the spam problem.

The fifth chapter revisits the spam solution activity from the third chapter, but this time with respect to the model presented in the fourth chapter. Solutions are grouped into categories based on how they interact with the Base Spam System. Then each category is examined in particular by placing it into the model, after which the effects of each placement are enumerated. In so doing, this fifth chapter evaluates spam solution types to date.

The final and sixth chapter continues to build on the previous chapters and considers how a more effective spam solution strategy might be structured. General points derived from earlier discussion are offered, along with an example strategy. The example strategy is not perfect nor is the only solution strategy, but simply an illustration of concepts presented. Again, the intended central contribution of this thesis is the model of the Base Spam System.

1.2 Reading this Thesis

This thesis is intended to be read in a linear fashion. However, depending on one’s knowledge of the spam problem, some sections could be skipped without significant loss of continuity. In particular, if one is familiar with the SMTP protocol and spam sending techniques, one could skip chapter two in its entirety. Similarly, if one is familiar with spam solution activity to date, one could skip chapter three in its entirety. However, even if one is familiar with the spam problem, it is nevertheless suggested that one read the rest of this chapter, as it introduces terms and concepts that are implicitly used in the model of the Base Spam System.

1.3 Why the Definition of Spam is Important

Most of the rest of this chapter is spent exploring possible definitions of spam and then settling on a useful definition for the purposes of this thesis and in general. This discussion may seem belabored, but is nevertheless important because:

1. There are many varying and conflicting definitions of spam currently in use. Making sense of this reality can illuminate important aspects of spam.
2. This thesis builds a model around spam, and thus this central term needs to be clearly defined for this model to be practically useful.
3. If one dominant definition of spam could emerge, activity currently spent arguing about what is spam and what is not spam could be refocused on more proactively solving the spam problem.
4. The choice of the definition of spam can shape the efficacy of spam solution activity, and thus have a pronounced impact on the timeliness of solving the spam problem.

These points will be brought out in the following sections. First, an ideal definition of spam is considered and discarded.
1.4 Difficulties Creating an Ideal Definition of Spam

An ideal definition of spam would have the following two properties:

1. **Parsimony.** The ideal definition would be precise enough such that the definition would declare sharp rules that could be used on any email message to tell if it were spam or not. That is, *ex ante* application of the definition would not result in a category of messages that require further investigation.

2. **Accuracy.** The ideal definition would be accurate such that if it declared a message as spam, it would indeed be spam.

This section highlights two difficulties in creating an ideal definition. To illustrate these difficulties, consider some of the arenas where spam is currently defined:

A. Spam as defined by individuals when dealing with the spam problem on a daily basis. Such definitions are personal and may turn on what email individuals consider unwanted and unsolicited. What is defined as spam for a particular individual depends on that individual's preferences, and is codified into his or her mental processes.

B. Spam as defined by spam content filters. Such definitions are used primarily by Internet Service Providers (ISPs) and Email Service Providers (ESPs) to filter spam messages for their customers. The definition in a particular instance is the algorithm used by a given spam content filter.

C. Spam as defined by email sent that was illegal to send as determined by existing laws within the jurisdictions from where and to which it was sent. Such definitions are used by governments, ISPs and ESPs to prosecute spam senders. A definition in this context would be the language of laws currently on the books in particular jurisdictions.

D. Spam as defined by email that is prohibited by private contract between email senders and their ISPs or ESPs. Such definitions are used by ISPs and ESPs to prohibit spam senders from using their resources to send spam. Definitions in this context would be the language of particular Acceptable Use Policies (AUPs) that were agreed to by particular users of ISP and ESP systems.

Definitions that arise from these arenas can easily conflict as to whether a particular message is spam or not. For example, a spam content filter (B) may declare a message as spam even though an individual (A) would not. Content filtering rules are based upon probabilistic models of individual preferences. But since these preferences vary by individual, particular declarations of spam can conflict with a particular individual's determination. In other words, the difficulty is what is spam to someone is not necessarily spam to someone else.

To achieve ideal accuracy, parsimony must be subjugated. That is, achieving ideal accuracy would have to involve individual decisions about what is spam in order to capture differences in individual preferences. In so doing, messages would have to be taken on a case by case basis, i.e. *ex post* application. Therefore, any definition achieving ideal accuracy

---

2 A spam content filter looks at the content of email messages, and declares messages as spam based upon that content.
would not be precise enough to draw bright lines around spam up front. In other words, the tradeoff between preciseness and accuracy is fundamental to a realistic definition of spam.

The second general difficulty on creating both a precise and accurate definition of spam is the ever-changing nature of what is generally considered spam. As more people have begun to use email, the market for spam has grown. In turn, email users have expressed strong preferences against spam, and as a result, prohibitions of spam sending in law (C) and private contract (D) have grown as well. As spam senders react to these somewhat precise definitions, they change tactics to evade them, and they no longer maintain accuracy.

In other words, what is generally considered as spam changes over time, and definitions need to change as well if they are to maintain the same levels of accuracy. The ideal definition is thus a perpetually moving target. To achieve both preciseness and accuracy, the rules that make the definition precise need to be constantly updated to reflect new spam tactics and constantly validated to achieve accuracy.

However, just because an ideal definition cannot be constructed does not mean that a slightly less than ideal definition cannot as well. Some accuracy or some preciseness or some of both will need to be sacrificed, but a definition of spam can still be useful. After all, even though what is generally considered spam is ever-changing and what people consider spam varies by individual, there exists a large subset of messages that people generally agree are spam and not spam. The next few sections will explore properties of realistic definitions.

1.5 Errors Made When Applying Realistic Definitions of Spam

The previous section highlighted general difficulties in pinning down a both precise and accurate definition of spam. In reality, a given definition will be less precise or less accurate than ideal, or both. When such a definition is applied to declare a particular message as spam or not, two errors can be made with regards to accuracy:

1. A message can be declared spam when it is actually not spam, i.e. a false positive.
2. A message can be declared as not spam when it is actually spam, i.e. a false negative.

First, note that the determination of whether one of these errors exists in a particular instance assumes that a given email can absolutely be declared spam or not. This assumption is by no means obviously valid given that this declaration is based on individual preferences and individuals have been known to be initially wrong when making such determinations, e.g. with email scams. However, one can suppose that an individual can ultimately tell whether a given message is spam or not according to their individual preferences.

3 Such an ideally accurate definition could be: whatever an individual ultimately considers spam, is spam, for that individual only. See the next footnote for a discussion of the word ultimately in this context.

4 To see this point, consider a particular individual's definition of what is spam is that which is unwanted and unsolicited. If the user initially declares a message is wanted but it later turns out to be unwanted, that would be a false negative. Yet to make the determination that the error occurred, i.e. that the email was really unwanted, the individual must have subsequently decided that the message was unwanted. Thus, even though their initial personal definition created a false negative, they were able to eventually absolutely declare the
Second, note that these errors are intrinsic to all definitions of spam applied in an ex ante fashion, excepting the ideal (unobtainable) definition. Therefore, any moderately precise definition of spam must wrestle with trading-off false positives with false negatives. Both have opportunity costs. False positives can result in missed actions that could have otherwise been taken and derived benefit to senders and recipients, and false negatives result in the familiar costs of sorting through spam messages for non-spam messages.

In the aggregate, both false positives and false negative also have macro effects on the email communication medium as a whole. These effects include the reliability, trustworthiness, and usability of email in general, and will be explored in later chapters. The tradeoff between false positives and false negatives is distinct from that between preciseness and accuracy. This distinction is explored in the next section.

1.6 Trading Preciseness for Accuracy When Creating Realistic Definitions of Spam

The tradeoff between preciseness and accuracy when creating realistic definitions of spam has been noted. Yet this tradeoff does not occur in a linear or otherwise predictable fashion. Differing definitions can vary wildly in their accuracy, but be similarly precise, and vice-versa.

Ideally, one would like to have a definition that is as precise and as accurate as possible. To see how the tradeoff is involved in practice, consider the precision side of a definition by itself. How precise a definition is involves how many messages one can say are spam or not in an ex ante process. This involves setting up rules to say whether a particular message is spam or not.

There are many messages for which precise rules can be made without diminishing accuracy, e.g. messages about sums of money from deceased Nigerian kings. Yet there are also many messages for which rules cannot be made. Here the choice is to be precise and mark them as spam (or not) even though the choice is not clear, or be accurate and deal with these messages on a case by case basis through individual investigation.

This tradeoff is thus distinct but related to the tradeoff between false positives and false negatives. Being precise with messages that are not clearly spam creates false positives and false negatives. Whether those messages are more wrongly declared spam or not spam is the tradeoff between false positives and false negatives. The ability to better distinguish spam from not spam thus allows you to be more precise without sacrificing accuracy.

1.7 Using Fraudulent Aspects of Messages When Creating Precise Rules

In the previous section, the tradeoff between preciseness and accuracy was explored. When constructing a definition of spam, there is also the question of how to be precise, i.e. what message as spam or not, just after significant deliberation, e.g. after acting on the message and receiving the results from their actions.
aspects of messages the precise rules will consider when declaring messages spam or not spam. The aspects chosen are a primary distinguishing factor between spam definitions.

One aspect that seems particularly useful is whether messages are using fraudulent tactics or not. Such tactics have evolved over time and continue to do so, but generally include tactics that intend to deceive or mislead recipients. Some specific tactics of this nature are outlined in the next chapter. Now consider the use of these tactics as a means to precisely declare messages as spam or not. To do so, let us briefly back up and consider why fraudulent tactics exist at all.

The spam problem arose and continues to exist because there are underlying economic market incentives that drive it. Spam reaches literally hundreds of millions of potential buyers, and some of those people do indeed respond. Additionally, the cost of reaching those people via spam is (currently) significantly less than the revenues generated.

It is a problem because while some people do indeed respond and thus in some sense want spam messages, the vast majority of people do not want the messages but nevertheless must deal with them. As mentioned, dealing with them takes time, and this time has opportunity cost. As a result, most people have become understandably concerned because they are bearing costs they do not want to bear.

The codification of anti-spam rules into laws and AUPs has grown in reaction to this concern. Just as there is a private economic interest in sending spam, there is also a private economic interest in alleviating concern to retain and capture customers. In addition, there is a public economic interest in getting back the aggregate opportunity costs lost when people deal with spam.

Definitions of spam within laws and AUPs are used to not only mark messages as spam, but also to identify spam senders in order to deny them access to spam sending resources. This latter use is driven by the attempt to stop the spam problem at its source, i.e. stop spam senders. In other words, if ISPs and ESPs could precisely and accurately identify spam senders on their networks, they could similarly precisely and accurately identify spam. The problem of identifying spam is thus reduced to the problem of identifying spam senders.

The basic way to identify spam senders is to trace spam sent to particular network users. As we will see in the next chapter, this is no easy task given the technical reality of email sending. Nevertheless, for spam senders to continue doing what is prohibited by laws and AUPs, they have had to engage in more and more fraudulent tactics to mislead and deceive recipients, ISPs, ESPs, and governments about their identities and intentions. Otherwise, they would be easily identified and quickly kicked off the network, and, perhaps also prosecuted.

Spam senders thus have significant incentives to engage in fraudulent behavior in order to avoid getting caught and to continue to make money from spam sending. On the other hand, non-spam senders, for the most part, have no reason to engage in fraudulent tactics

---

5 ISPs and ESPs must also bear the cost of transferring these messages, and in some cases, dealing with them on behalf of their customers.
when sending email. Therefore, if spam is at least partially defined with regards to the use of fraudulent tactics, such a definition may usefully isolate spam senders in such a way that they can be more easily prosecuted by governments and shut off by ISPs and ESPs. We now consider this point in an attempt to construct a useful definition of spam for the purposes of the rest of this thesis.

1.8 Constructing a Useful Definition of Spam

To summarize the previous sections, when constructing a useful definition of spam, the following things should be considered:
1. The tradeoff between preciseness and accuracy.
2. The tradeoff between false positives and false negatives.
3. The stratification of spam senders and non-spam senders through highlighting the use of fraudulent tactics.

Now, a top-down approach is taken to arrive at a particular definition for use throughout the rest of this thesis.

Consider the following two properties of a given email:
1. Whether the email uses fraudulent tactics or not; and
2. Whether the email contains a valid opt-out mechanism.

The latter property refers to a process whereby the recipient of an email can contact the sender of the email to request to receive no more emails from the sender. These properties are represented in the following table (Table 1).

<table>
<thead>
<tr>
<th>Fraudulent/ Opt-Out</th>
<th>Yes (Opt-out Does Work)</th>
<th>No (Opt-out Does Not Work)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes (Fraudulent)</td>
<td>Spam</td>
<td>Spam</td>
</tr>
<tr>
<td>No (Not Fraudulent)</td>
<td>Not Spam?</td>
<td>Spam</td>
</tr>
</tbody>
</table>

There is general agreement that fraudulent email is spam irregardless of whether a valid opt-out mechanism exists or not. Additionally, there is also general agreement that non-fraudulent email that nevertheless does not have a valid opt-out mechanism is also spam because the recipient cannot engage in a meaningful process to stop future messages. Therefore, I believe that declaring these categories as spam is not particularly controversial and thus those declarations will be accepted as part of the definition of spam for the purposes of the rest of this thesis.

---

6 An exception may be the desire for anonymous email communication. However, to the extent that this market exists, re-mailing services could be expanded to fill this market as needed.
7 This property applies to all email, including non-commercial and personal email. In these cases, “opting-out” would include the ability to reply and ask to be removed from further communication from that person or organization.
8 Note that still, from previous discussion, declaring all fraudulent email as spam can lead to false positives depending on individual preferences on what is spam or not. For instance, an anonymous email (achieved by using fraudulent tactics) that is nevertheless wanted by a given recipient could be viewed as a false positive under a definition that considers all fraudulent email as spam.
9 Again (see last footnote), while there is general agreement, such a precise definition can lead to false positives.
Controversy arises in the last category, however. This category is email that does not use fraudulent tactics and does establish a valid opt-out mechanism. As such, consider further dividing this sub-category along the following two sub-properties:

1. Whether the email is commercial in nature.
2. Whether the email was solicited.

These sub-properties are represented in the following table (Table 2).

| Table 2 Sub-categories of Email that is Not Fraudulent and has Working Opt-out |
|----------------------------------------|--------|-------------|
| Commercial/Solicited                  | Solicited | Unsolicited |
| Commercial                             | Not Spam | Spam?       |
| Non-Commercial                         | Not Spam | Not Spam?   |

Solicited email in the sense that the recipient knowingly took some action to receive the email and expects it to come is generally agreed as not spam. Remember that in this table, all of these messages contain valid opt-out mechanisms such that recipients can stop future messages from occurring if desired.

First, note that often the difference between unsolicited and solicited is hard to prove in practice. People commonly forget what they signed up for, and they also do not know a priori what to expect when explicitly allow companies’ “partners” to send them messages. Nevertheless, if we assume email is truly unsolicited, both the commercial and non-commercial sub-categories of this type of email are questionable as to whether they are absolutely spam or not.

Non-commercial unsolicited email includes much email from friends and family, and therefore is generally thought of as not spam. However, more borderline cases can exist such as unsolicited email related to political or religious causes. The commercial side contains email that is often considered spam by individuals. However, it can also include things like someone asking you personally for a job.

Given that there are clearly both spam and not spam in these sub-categories, there might be a tendency to further subdivide them using more properties, such as whether email was sent in bulk. And that is certainly what some definitions of spam do. However, each property added means a potential fight over whether that property truly applies or not to a given message.

A simpler approach would be to back up and declare the entire second table as not spam, and stick to a definition as outlined in the first table (Table 1). The United States government has essentially taken this approach. Email that does not use fraudulent tactics and includes a valid opt-out mechanism is considered not spam. This thesis also adopts this approach, i.e. only the three boxes labeled ‘Spam’ in Table 1 are indeed spam. The next section explores specific reasons for this adoption.

---

10 The only significant difference is that they only consider commercial email, as defined by statute (CAN-SPAM Act of 2003, Sec. 3 (2)), to have the possibility of being spam. In other words, non-commercial email is categorically not spam even if it does not have a valid opt-out mechanism.
1.9 What is Spam?

Spam, for the purposes of this thesis, is email that uses fraudulent tactics or does not contain a valid opt-out mechanism. There are several arguments for why this is a useful definition to adopt.

First, this definition is simple. A set of fraudulent tactics can be enumerated into an extended definition, and as new fraudulent tactics evolve over time, they can be included as well. The details of a valid opt-out mechanism can similarly be enumerated into an extended definition.

Second, agreement on this definition can focus spam solution activity. In particular, solutions can then focus on creating algorithms to detect fraudulent tactics and valid opt-out mechanisms ex ante. To the extent that this is not possible given the technical realities of email protocols, those protocols can be augmented to allow for ex ante application.

Third, agreement on the definition and refocusing of spam solution activity would further stratify spam senders from non-spam senders. That is, non-spam senders would have more incentives to not use fraudulent tactics and to have valid opt-out mechanisms. Spam senders would similarly be forced to reveal their identities and deal with customer complaints or risk being shut off by ISPs and ESPs and prosecution.

Fourth, this definition largely captures spam today. A Federal Trade Commission (FTC) study in 2003 found that about two thirds of spam messages contained fraudulent content of one form or another (Federal Trade Commission, 2003). If the definition was agreed to in some critical fashion, this percentage would be expected to rise as the stratification of non-spam senders from spam senders continued. In other words, the act of accepting this definition would in and of itself tend to capture more of spam in the future.

There are drawbacks, however. First, this definition is not as precise as other definitions. That is, applying it today to a large subset of messages would require some case by case investigation, or a lot of false positives or negatives. However, this drawback can largely be mitigated by refocusing spam solution activity to create the technical means for ex ante application of the definition. More on this will be explored in subsequent chapters.

Second, this definition allows unsolicited email from a given legal entity, once. This first email may currently be considered spam by many recipients, and thus may result in many false negatives. However, individual preferences could change over time given agreement on this definition. That is, if email recipients understood that this first email was allowed and believed that opt-out mechanisms worked, many of these false negatives would disappear as individual ideas about spam changed.

Furthermore, if too many false negatives were occurring, the definition could easily be expanded to include more sub-categories, like was started to do in the previous section. However, it is not at all obvious that such further distinctions are needed, so until that is proven, the simpler definition seems the better choice.
Now that a definition of spam has been adopted, the general nature of the spam problem that will be analyzed in subsequent chapters is examined. As long as the costs of sending spam remain significantly lower than the revenues generated from responses, there will be a significant market opportunity to send spam because of the profit opportunity. Yet despite this large opportunity, only a handful of people send significant amounts of spam—fewer than 200 spam operations send about 90% of the spam.

This striking proposition is made by Spamhaus, a vigilante group that tracks spam operations (Spamhaus, 2004). Note that this does not mean that only two hundred products or services are sold via spam. Some of these operations contract out their services to other businesses; in other words, they have specialized in sending spam in the global marketplace.

In the aggregate, these operations are sending a lot of spam. Spam estimators claim that over 12 billion spam messages are sent each day, and now account for over 60% of all messages.\(^1\) Note that the ratio of spam to non-spam for actual recipients is not necessarily the same percentage, because many of the spam messages are filtered out in one way or another before they reach inboxes.

For recipients, this latter metric, i.e. the ratio of spam to total email received, is the real barometer of the spam problem. For the purposes of the rest of this thesis, this metric is termed the spam ratio. If the spam ratio is small, e.g. 0.05 (5%), then spam is empirically trivial to deal with because only 1 in 20 messages is spam. The spam problem has arisen as the spam ratio has increased, which it has significantly in recent years.\(^2\)

As this has happened, the problem has become manifest through the visible costs of dealing with false positives and false negatives. As mentioned earlier, in the aggregate, these costs can lead to loss in reliability, trustworthiness, and usability of email as a communications medium in general. These effects are examined in detail in subsequent chapters through the model of the Base Spam System.

In this chapter, spam and the spam problem have been defined. In so doing, various aspects of the spam problem were mentioned, including the use of fraudulent tactics by spam senders and the costs of false positives and false negatives. In the next chapter, the technical process of sending email is enumerated, and the exploitation of this process is explored.

---

\(^1\) Brightmail, self-pronounced “Anti-Spam” leader maintains spam statistics page at and currently claims (as of April 2004), that 64% of all email transmitted is spam (Brightmail, 2004). Spam Filter Review, a Web site that reviews spam filter software, maintains broader, but slightly older statistics and report that in for 2003, on average, 31 billion emails are sent every day, 12.4 billion of which are spam, or approximately 40% (Spam Filter Review, 2004). Also see SpamCon Foundation, 2004.

\(^2\) This ratio is harder to monitor because of the filtering problem mentioned. If assumed roughly tied to the ratio of wanted email to all email that is sent, Brightmail reports this ratio has risen from 48% to 64% since May 2003 (Brightmail, 2004). Some isolated examples of actual spam ratios do exist though; these include an AOL account (TESP.com, 2004) and a user in the Netherlands (Wouters, 2004).
2 Sending Email

In the first chapter, spam and the spam problem were defined. In this chapter, the technical process of sending email is enumerated, and the general exploitation of this process is explored. Today, almost all Internet users have one or more email accounts that they use to regularly communicate both personally and commercially (Pew Internet & American Life Project, 2004). This huge email user base is the underlying driver of spam—the same underlying driver that drives junk mail and junk faxes—an opportunity to reach people. For email, this opportunity is especially significant because the cost of sending email is extremely low to the sender.

Moreover, email has been increasingly used for commercial use, but the basic email protocols were not designed with commercial use in mind. Over the years, end-user email applications have adapted to enable more effective commercial use, e.g. by displaying in-line product graphics and order forms. Yet even as this expansion has taken place, the basic email protocols have not changed significantly. I will now introduce these protocols, and explain why the email system they comprise has enabled the spam problem to take its current form.

2.1 Basic Email Protocols

The fundamentals of the email technology that we use today were developed before the spam problem was a problem, and even before the widespread use of the Internet.13 There are three main protocols in use to send and receive basic email. The standard way to send email is to use the Simple Mail Transfer Protocol (SMTP), which was officially proposed in 1982 (Postel, 1982). To receive email, users commonly use either the Post Office Protocol (POP), which was officially proposed in 1984 (Reynolds, 1984), or the Internet Message Access Protocol, which was officially proposed in 1996 (Crispin, 1996). These protocols have been updated, but in so doing, they were not significantly modified with respect to the spam problem.14

Consider the sending of an email from a sender to a recipient (see Figure 1). The sender commonly composes the email in an email “client” (end-user application). Email clients are the pieces of software that email users commonly associate with email, and include such popular desktop applications as Microsoft Outlook and Eudora. These clients typically have the ability to use all three main protocols. When sending an email, the client transmits the email message to the Sender's Email Server via SMTP. (An email server generally is a computer that transmits email messages through the Internet to other email servers, and can make particular email messages available to users for download as appropriate.) This email

---

13 Some say the first email considered spam by some people was sent in 1978 (Templeton, 2004), while others point to a message in 1994 (Everett-Church, 1999). As such, when exactly the “spam problem” became a real problem is certainly debatable, but was definitely by 1996 (Windigo, 1996). When “the widespread use of the Internet” occurred is also debatable, but again, 1996 is a decent first-order estimate (Ratnatunga, 2004).

14 The latest major revision of SMTP was in 2001 (Klensin, 2001); the latest major revision of POP was in 1996 (Myers et al., 1996); and the latest major revision of IMAP was in 2003 (Crispin, 2003)
server is usually associated with and maintained by the sender’s Internet Service Provider (ISP) or organization.

Figure 1 Email Protocols in Use

Once the message is on the Sender’s Email Server, that server transmits the message to what it believes to be the Recipient’s Email Server via SMTP in the same fashion the sender’s email client transmitted the message to it. The email server that the Sender’s Email Server transmits the message to, however, may in fact not be the Recipient’s Email Server if the recipient or his/her ISP or organization has set up some sort of email forwarding system. For example, many people maintain vanity email addresses, which forward to other email accounts. If there are Intermediate Email Servers, the message gets transmitted in the same fashion via SMTP until it reaches the final Recipient’s Email Server.

Once the Recipient’s Email Server has the message, it deposits the message in a place designated for the recipient. The recipient can then pick up the message when desired, using either the POP or IMAP protocols. The primary difference between these two receiving protocols is that with IMAP, the message remains on the server, whereas with POP, the message is removed from the server. Users that use Webmail clients, e.g. Hotmail, would
not necessarily use either POP or IMAP, but would connect directly with the last server via a Web application.

Amazingly, these protocols have continued to perform effectively as described while email traffic has increased by many orders of magnitude. Nevertheless, the Internet is a very different place now (in both size and nature) from the early 1980s, and if the protocol designers could have seen this future, they might have designed the protocols differently. With regards to the spam problem in particular, the significant architectural issues lie within SMTP. POP and IMAP are just used for receiving email that already exists, and therefore, are not particularly relevant to the spam problem. To understand technically how spam and spam solutions work, it is necessary to understand how basic SMTP works. As such, I will now explain SMTP in more detail.

2.2 SMTP Basics

Consider the first step in the above discussion and picture, namely the transmission of an email message from the Sender's Email Client to the Sender's Email Server via SMTP. As noted, the other SMTP transmissions between the email servers occur in the same fashion as this first step, and thus we can concentrate on only this first step (see Figure 2).

Figure 2 SMTP In Use

1. (connect) (ACK)
2. HELO (ACK)
3. MAIL FROM (ACK)
4. RCPT TO (ACK)
5. DATA (ACK)
6. QUIT (close)

Sending a standard email via SMTP is a six step process, including connecting and disconnecting from the email server. I will outline the steps, and then present a detailed example. First, a connection is made to the server, which can be done through a variety of means (connect), and the server sends back an acknowledgement (ACK). Second, the HELO command is sent identifying the sender's computer, followed by an acknowledgement from the server. Third, the MAIL FROM command is sent identifying the sender, followed by an acknowledgement from the server. Fourth, the RCPT TO command is sent identifying the receiver, followed by an acknowledgement from the server. Fifth, the DATA command is sent including the message itself, followed by a final
acknowledgement from the server. Finally, the QUIT command is sent, followed by the server closing the connection (close).

To make this process more clear, I will now enumerate a real life example. Consider the act of sending an email manually using the basic SMTP process from the figure. Not many people actually send email this way. However, this is the same process most email clients go through automatically when people hit the send buttons in their email clients, and it is instructive to illustrating major spam techniques. What I sent to the email server appears in **bold**, and what the server sent back to me is in *italics*.

1. First, I connected to the email server via telnet (connect), which immediately acknowledged my connection (ACK) by identifying itself.

   220 MIT.EDU ESMTP Sendmail (no collect or third party calls) at Fri, 2 Jul 2004 16:56:19 -0400 (EDT)

2. Next, I began the standard SMTP exchange (Yahoo, 2004). First, I identified myself with the HELO command followed by my domain name. The server responded affirmatively (ACK), which you can tell by the leading ‘250’ response.

   HELO MUCKLEY-THREE-NINETY-FOUR.MIT.EDU
   250 fort-point-station.mit.edu Hello MUCKLEY-THREE-NINETY-FOUR.MIT.EDU [18.172.6.138], pleased to meet you

3. Then I told the server where this message came from using the MAIL FROM command. The server acknowledges affirmatively (ACK) via the ‘250’ response.

   MAIL FROM: <YEGG@ALUM.MIT.EDU>
   250 2.1.0 <YEGG@ALUM.MIT.EDU>... Sender ok

4. Then I told the server where this message is going using the RCPT TO command. The server acknowledges affirmatively (ACK) via the ‘250’ response.

   RCPT TO: <YEGG@MIT.EDU>
   250 2.1.5 <YEGG@MIT.EDU>... Recipient ok

5. Next, I actually specified the email to be sent using the DATA command. The email is broken into two sections, the header and the body. The header contains meta-information about the email, such as the subject. Each header field has a name, e.g. ‘Subject’, and then an associated value, e.g. ‘TEST SUBJECT’, separated by a colon and a space. Then one blank line separates the header from the body. Finally, a single dot on a line by itself signifies the end of the message. The server then acknowledges affirmatively (ACK), again via the ‘250’ response.

   DATA
   354 Enter mail, end with "." on a line by itself
   FROM: YEGG@ALUM.MIT.EDU
   TO: YEGG@MIT.EDU

A Systems Analysis of the Spam Problem
SUBJECT: TEST SUBJECT

TEST BODY

250 2.0.0 i62K1VRg018045 Message accepted for delivery

6. Finally, I said goodbye using the QUIT command, and I assumed the mail will be
delivered successfully, given all the affirmative responses. The server closes the
connection (close).

QUIT

221 2.0.0 fort-point-station.mit.edu closing connection

2.3 Email Format

After sending the email above, I had my email client, Microsoft Office Outlook 2003,
receive the mail using the Post Office Protocol (POP) (Yahoo, 2004). After receiving the
email, I went to View→Options, and copied the full headers of the email received, and
past them below.

Status: U
Return-Path: <YEGG@ALUM.MIT.EDU>
Received: from alum-2.mit.edu ([18.7.21.145])
    by james.mail.atl.earthlink.net (EarthLink SMTP Server) with ESMTP id 1bGwdk7e83Nl3r10
    for <yegg@mindspring.com>; Fri, 2 Jul 2004 18:08:58 -0400 (EDT)
Received: from fort-point-station.mit.edu (FORT-POINT-STATION.MIT.EDU [18.7.7.76])
    by alum-2.mit.edu ([8.12.8p2/8.12.8] with ESMTP id i62M8w24025072
    for <yegg@ALUM.MIT.EDU>; Fri, 2 Jul 2004 18:08:58 -0400 (EDT)
Received: from MUCKLEY-THREE-NINETY-FOUR.MIT.EDU (MUCKLEY-THREE-NINETY-FOUR.MIT.EDU [18.172.6.138])
    by fort-point-station.mit.edu (8.12.4/8.9.2) with SMTP id i62M7POW023041
    for <YEGG@MIT.EDU>; Fri, 2 Jul 2004 18:08:35 -0400 (EDT)
Date: Fri, 2 Jul 2004 18:07:25 -0400 (EDT)
Message-Id: <200407022208.i62M7POW023041@fort-point-station.mit.edu>
FROM: YEGG@ALUM.MIT.EDU
TO: YEGG@MIT.EDU
SUBJECT: TEST SUBJECT
X-EFL-Spamscore: 59%
X-Spam-Flag: NO
X-ELNK-AV: 0

The message format in general is governed by a proposal related to SMTP (Resnick, 2001).
The first thing to note from this example is that this header is much larger than the header
that I sent. Recall that I only sent 'From', 'To', and 'Subject' header lines, but yet there are
many other headers lines here, including several 'Received' lines, an 'X-Spam-Flag' line, etc.
Additionally, hardly any of these extra header lines are displayed in my inbox view. In fact,
all I see from the headers above in my inbox view is the following.

From: YEGG@ALUM.MIT.EDU
To: YEGG@MIT.EDU
Subject: TEST SUBJECT
The extra headers were added in the process of sending the email, and are used by email clients and servers for various purposes. Certain headers have particular universal meanings. For instance, each email server that handles the message is supposed to add a 'Received' header line that specifies who that server received the message from and at what time. You can see that my message to YEGG@MIT.EDU was actually forwarded to YEGG@ALUM.MIT.EDU, and then forwarded to YEGG@MINDSPRING.COM. Thus after I transmitted the message to the MIT SMTP server, it transmitted the message to the MIT Alum SMTP server, which then transmitted the message to the Earthlink SMTP server, and each server added a 'Received' header line in the process.\footnote{Earthlink, Inc. bought MindSpring Enterprises in 1999.}

Other non-standard fields can be added as needed, and are usually preceded by 'X-'. Email servers are supposed to simply pass these header lines through. In the email that I received, you can see that three such lines were added (X-EFL-Spamscore, X-Spam-Flag, X-ELNK-AV), which have to do with spam and anti-virus programs.

\subsection{SMTP Exploitation}

The SMTP exchange described in the previous sections is often exploited by spam senders. This section describes a primary way in which such exploitation is done. Consider the following example exchange conducted by me with the MIT SMTP server. Again, the \textbf{bold} signifies what I sent to the email server, and the \textit{italics} signify what the server sent back to me.

\begin{verbatim}
220 MIT.EDU ESMTP Sendmail (no collect or third party calls) at Fri, 2 Jul 2004 16:56:19 -0400 (EDT)
HELO WHITEHOUSE.GOV
250 fort-point-station.mit.edu Hello MUCKLEY-THREE-NINETY-FOUR-MIT.EDU [18.172.6.138], pleased to meet you
MAIL FROM: <GWBUSH@WHITEHOUSE.GOV>
250 2.1.0 <GWBUSH@WHITEHOUSE.GOV>... Sender ok
RCPT TO: <YEGG@MIT.EDU>
250 2.1.5 <YEGG@MIT.EDU>... Recipient ok
DATA
354 Enter mail, end with \".\" on a line by itself
TO: YEGG@MIT.EDU
X-WHATEVER: WHATEVER
.
250 2.0.0 62KajfW029811 Message accepted for delivery
QUIT
221 2.0.0 fort-point-station.mit.edu closing connection
\end{verbatim}

I have apparently sent a message purportedly from the White House, and specifically from an email address that might be associated with the US President. This does not have to be done manually, and can actually be done easily in most email clients.\footnote{In Microsoft Outlook, for example, when one composes a message, one can go to View $\rightarrow$ From Field, and then type any address next to the From box.} Again, the following is what I see as the headers upon receipt.
From this example, you can see that everything sent to the email server can be easily exploited (forged). The one notable exception is the 'To' header field, which is needed to actually deliver the email to the intended recipient. Additionally, only the 'To' field is essential, and in this example I only sent the 'To' field and an 'X-WHATEVER' field, which I just made up and which was passed through intact as expected. I could similarly make up and include other 'X-' fields to make a particular message seem like it came from certain email clients, such as Outlook or Eudora or Hotmail. To do so, I would just add the extra header lines that those email clients usually add, and in most cases there is really no way for the recipient to tell the difference.

Furthermore, I could add my own 'Received' headers to pretend that the email was received and relayed by various email servers throughout the Internet. Note, however, that the final email server, and any email servers in-between, can always add their own additional fields to whatever I send, including their own 'Received' header lines. In other words, I can make my message seem like it passed through various hops or came from a particular email client, but I can also do nothing to prevent MIT from adding its own headers. For instance, the 'X-EFL-Spamscore' line in this example was not added by me.

Finally, note that the SMTP exploitation outlined in this section is the exploitation of the SMTP exchange through the submission of information that is fraudulent, i.e. not representative of the true sender. Such exploitation is distinct from the act of accessing SMTP servers that one is not authorized to access. That is, many SMTP servers are only open to certain computers, often those that are maintained by the same organizations that maintain the SMTP servers. For example, an ISP might only let its customers send email through its SMTP servers.

This process of authenticating and authorizing certain computers or users to use certain SMTP servers is called SMTP Authentication. The point of this process is to authenticate that that senders connecting to the email servers are who they say they are and that they are authorized to connect (Myers, 1999). However, once authenticated and authorized, nothing
prevents the SMTP exploitation as outlined in this section to take place. For example, MIT uses a form of SMTP Authentication, but once I connected to the email server successfully, I could engage in the above exploitive exchange notwithstanding.

2.5 IP Address Basics

In the previous section, SMTP exploitation was explored. Given the possibility of such exploitation, it follows that email exchanges are not particularly "trustable." In fact, the only inherently trustable part of the exchange is that to make an exchange, the sender has to connect to the server from somewhere. That is, by connecting to the server, the sender has to actually send information to it through the Internet, and by so doing, expose an IP address to it. IP addresses are numerical codes that identify devices on the Internet and are used for the very purpose of exchanging information with other devices on the Internet (Wikipedia, 2004).

In the example exchange in the previous section, the IP address used to connect to the server is displayed after the HELO line. The server printed it in its response, in brackets [18.172.6.138]. Servers do not have to display the sender IP address in brackets in this manner, but they usually include this information in the 'Received' headers they add to emails, and thus, by looking at the first received header in a given email (chronologically), one can usually reliably determine the IP address from which the email originated. Even if the server does not display the address, it is aware of the address because it needs to know it to communicate effectively with the sender. Additionally, this information can and often is logged in the background for future reference.

Given a reliable IP address, one can begin to reliably identify the sender. This identification stems from the fact that one cannot simply make up an IP address and then hope to connect to the Internet in a useful manner. On the contrary, the IP address is the essential piece of information needed to route information, like a mailing address in the real world. Just like the postal service issues mailing addresses, the connectors to the Internet, e.g. ISPs, issue IP addresses.

In other words, given a working IP address, by definition, allows one to send information to that device connected to the Internet. The route used to send that information can be analyzed, much like tracking a regular snail mail through various postal service centers.\(^\text{18}\) At

\(^{17}\) From this point, one should not conclude that SMTP Authentication is not useful, however. In the scenario explored, MIT would have a better idea who I am from the SMTP Authentication process, and thus could potentially more usefully act to prevent further SMTP exploitation. For example, they could prevent my future access.

\(^{18}\) For example, using the popular 'traceroute' program, I can “ask” routers that are in the route between MIT and my home computer to list their IP addresses. The following is the output returned.

```
traceroute to 24.218.138.161 (24.218.138.161), 30 hops max, 40 byte packets
 1 W92-RTR-1-W92SRV16.MIT.EDU (18.7.16.1) 0.652 ms 0.729 ms 0.401 ms
 2 EXTERNAL-RTR-2-BACKBONE.MIT.EDU (18.168.0.27) 0.479 ms 0.467 ms 0.480 ms
 3 24.218.0.193 (24.218.0.193) 0.594 ms 0.631 ms 0.590 ms
 4 bar02-p4-0.ndhmhe1.ma.attbb.net (24.91.0.158) 0.952 ms 0.862 ms 0.834 ms
 5 65.96.1.154 (65.96.1.154) 0.946 ms 0.966 ms 0.974 ms
```
some point when tracing a route to a particular IP address, a router is reached from the
issuing party of that IP address, be it an ISP, country, educational institution, etc. Then one
can go to that entity and ask it to reveal more information associated with that particular IP
address. Even if that is not possible, one can often pinpoint the IP address around a rather
tight geographical location.  

2.6  **IP Address Exploitation**

From the previous sections, we conclude that the IP address is the only relatively trustable
part of the SMTP exchange. As such, it has been used considerably in various attempted
and proposed spam solutions to date. These solutions are outlined in the following chapter.

As noted, one way the IP address can be used is in an attempt to trace the sender. To the
extent that this is possible, the IP address is obviously useful in enforcement efforts. It is
important to note, however, that an IP address identifies just a *device*, and not a *person* that
was actually operating the device at some time. As such, this relationship between person
and device can be exploited by using IP addresses that are not easily traceable beyond the
device to the person operating the device. Such IP Address Exploitation can be conducted
through a variety of means, some of which are briefly noted below.

- **Open Proxies**
  Open proxies are servers that enable people to connect remotely to a network, and
then communicate to the outside world as if they were actually located within the
network that they are connecting through (Wikipedia, 2004). For example, a
connection can be made from a home computer to a work proxy server by
connecting directly by dial-up modem or through the Internet via telnet or other
mechanisms. Once connected, the home computer can then access other machines
around the Internet through the proxy. The proxy acts as an intermediary, relaying
information from the remote servers to the home computer. In the process, it only
exposes its IP address to the remote servers and not the IP address of the home
computer.

---

6  c-24-218-191-178.ne.client2.attbi.com (24.218.191.178) 1.010 ms 1.074 ms 0.951 ms
7  ***
8  ***
9  ***

19 Several companies specialize in mapping IP addresses to geographical locations and other information. For
example, Maxmind (MaxMind, 2004) sells several such databases, and Visualware sells a product called
VisualRoute (Visualware, 2004), which is a graphical route tracing program that maps the actual route locations
across the globe.

20 The word *relatively* is used because it is possible to forge or “spoof” an IP address, but doing so involves
significant guess work and is extremely difficult with newer routers that have anti-spoofing technology built in
(Tanase, 2003). As such, spoofing IP addresses when conducting an SMTP exchange is not practical because
the sender can not see the exchange, and thus is unable to see if it is proceeding successfully or not. Recall that
there is not just one step in the SMTP exchange, but six, all with acknowledgements from the server. Spoofing
would thus require guessing correctly when steps have completed, and hoping that all acknowledgements are
affirmative.
• **Open Relays**
Open Relays are email servers that will accept mail for any domain, and then relay the mail not for their internal domains to the appropriate parties (Open Relay DataBase, 2004). In and of themselves, open relays do not completely mask the original IP address, which should be added in the ‘Received’ header of messages as discussed previously. However, they make it harder to stop certain IP addresses from sending email because those IP addresses are no longer connecting directly to the recipients’ SMTP servers. Instead, the open relay connects to those servers on behalf of the original IP address. Thus, in order to block certain IP addresses from originating mail to a particular server, the recipients’ SMTP servers need to examine the headers of the messages (content), which is significantly slower than just blocking certain IP addresses from connecting. Furthermore, when open proxies are connected to open relays, one can originate an email from the network of the open proxy and then send email through another network’s open relay, leaving no trace of the original IP address on the email itself.

• **Free Web-based Email Accounts**
When free Web-based email accounts are used to send messages, the server’s IP address connects to recipients’ SMTP servers on behalf of the person connecting to the Web server. In other words, the IP address of the user connecting to the Web-based email account is not included in the ‘Received’ header, but instead the IP address of the server is included. Some, but certainly not all free Web-based email services, have begun to add the user’s IP address to the email header via various header fields, such as the ‘X-Originating-IP’ field. Again, however, this security mechanism is undermined if the user is connecting to their Web-based email account via an open proxy. In that case, the so-called originating IP address that is added to the email header is not really that of the sender, but instead that of the open proxy. Also note that non-free Web-based email accounts are also dangerous in the same fashion, but are less dangerous because they are harder to obtain since valid payment information must usually be exchanged before use.

• **Hijacked Machines – Zombies**
Another common way people use IP addresses that are not traceable to themselves is to simply use others’ machines. This can either be done physically by breaking and entering or using a shared computer lab or through software by breaking into the machines through the network and then operating them remotely. Recently the latter tactic has been accomplished via viruses, not coincidentally that have been spread via email. Note that through viruses and other available tools, breaking into machines through the network does not require great technical expertise in all instances. (The machines compromised by viruses that send email on behalf of spammers are often referred to as “zombie” machines.)

There are certainly other ways to conduct IP address exploitation, and these ways may certainly evolve further over time into still new ways. However, the general notion of attempting to exploit the email system to hide ones identity will remain the same.
This chapter has introduced basic email sending processes, and shown some ways in which those processes they can be exploited. Exploitation in this context is attempts to fraudulently identify the senders of email messages or the devices from which they originated. Some of these exploitation techniques, those mentioned in the text, are shown in the following table (Table 3).

<table>
<thead>
<tr>
<th>Technique</th>
<th>Level of Difficulty to Do and Not Easily Get Caught (1-Easy to 5-Hard)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forge email headers</td>
<td>1</td>
</tr>
<tr>
<td>Manually Use Web-based Email Accounts</td>
<td>2</td>
</tr>
<tr>
<td>Find and use Open Proxies</td>
<td>3</td>
</tr>
<tr>
<td>Find and use Open Relays</td>
<td>3</td>
</tr>
<tr>
<td>Sign up Fraudulently with ISPs</td>
<td>3</td>
</tr>
<tr>
<td>Automatically Use Web-based Email Accounts</td>
<td>4</td>
</tr>
<tr>
<td>Use Existing Tools to Break Into Machines</td>
<td>4</td>
</tr>
<tr>
<td>Create New Tools to Break Into Machines</td>
<td>5</td>
</tr>
</tbody>
</table>

The second column in the table considers how difficult the techniques are to conduct without getting caught. These levels of difficulties are of course estimates, and may vary from one situation to the next. Additionally, they may change significantly over time. I have included them to indicate difficulty in a relative sense across the techniques.

The next chapter reviews spam solution activity, and explains many attempted and proposed solutions to date. Some of these solutions attempt to make the exploitation techniques explored in this chapter harder to do without easily getting caught.
3 Review of Spam Solution Activity

The previous two chapters have attempted to partially characterize the spam problem. This chapter builds on that characterization by reviewing spam solution activity to date. Please note that this chapter is not intended to be an evaluation of solutions, but instead a compendium of the types of attempted and proposed solutions. In other words, this chapter is intended to explain how each type basically works so that each type can be considered in more detail in relation to the Base Spam System model in the fifth chapter.

3.1 Overview of Major Spam Solution Activity

The spam problem really impacts everyone that uses email, which includes most people that use the Internet. As such, many groups and individuals across the globe are actively trying to attack this problem from a variety of angles.

The following table outlines the major spam solution activity (Table 4). Each row represents a logically different solution type. These 21 different types are grouped into six categories based on similarities in how they interact with the spam problem. In subsequent sections, each category and type will be basically explained. This explanation will be enhanced in subsequent chapters after the Base Spam System model is introduced in the next chapter.

<table>
<thead>
<tr>
<th>Category</th>
<th>Name</th>
<th>Leading Program/Proposal</th>
<th>Leading Organization</th>
<th>Implemented (1 No - 5 Yes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content Filtering</td>
<td>Rule-based Filtering</td>
<td>SpamAssassin, Brightmail Anti-Spam ISP</td>
<td>The Apache Software Foundation, Brightmail, Inc.</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Collaborative Filtering</td>
<td>SpamNet</td>
<td>Cloudmark, Inc.</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Client White Lists</td>
<td>Embedded in Client Software</td>
<td>Email Client Providers</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Challenge/Response</td>
<td>Earthlink spamBlocker</td>
<td>Earthlink, Inc.</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Client Black Lists</td>
<td>Embedded in Client Software</td>
<td>Email Client Providers</td>
<td>5</td>
</tr>
<tr>
<td>Access Filtering</td>
<td>Best Practices for ISPs, Businesses, &amp; Consumers</td>
<td>ASTA Technology and Policy Proposal</td>
<td>ASTA</td>
<td>2</td>
</tr>
<tr>
<td>Access Filtering</td>
<td>Postage Costs</td>
<td>Trusted-Class Email</td>
<td>Goodmail Systems, Inc.</td>
<td>1</td>
</tr>
</tbody>
</table>
The Implementation column (fifth) warrants some explanation. I have rated each solution type (row) on a one to five scale on how widely deployed and available is the particular program or proposal. This scale is as follows:

1. Not being actively pursued yet—no major proposals or significant reference implementations.
2. Being actively pursued—major proposal and/or significant reference implementation exists.
3. Being used in a minor way—one or more minor release implementations exist.
4. Being used in a major way—at least one major release implementation exists.
5. Widely deployed—many major release implementations exist.

Again, the following sections will describe in more detail the activity in each of the categories (column 1) listed in the table. In this chapter, I will not, however, considerably explore the

| Access Filtering | Computational Costs | n/a | Microsoft | 1 |
| Access Filtering | Server White Lists | Bonded Sender | IronPort Systems, Inc. | 5 |
| Access Filtering | Server Black Lists | RBL, ORDB | MAPS, Inc., ORDB.org | 5 |
| Access Filtering | Do Not Email Registry | CAN-SPAM Act of 2003 | Federal Trade Commission | 2 |
| Access Filtering | Reverse DNS | AOL | America Online, Inc. | 5 |
| Identifiability | IP Address Authentication | Sender ID | Microsoft/PObox.com | 4 |
| Identifiability | Content Signing | DomainKeys | Yahoo! | 3 |
| Regulation | State Spam Protection Laws | Washington State | Washington State | 4 |
| Regulation | Acceptable Use Policies | AUPs | ISPs and ESPs | 5 |
| Enforcement Capacity | Lawsuits | AOL Legal Department | America Online, Inc. | 5 |
| Enforcement Capacity | Bounty | CAN-SPAM Act of 2003 | FTC | 2 |
| Enforcement Capacity | Vigilante Efforts | ROSKO | Spamhaus | 5 |
success or failure of the measures, or their positive and negative effects. That discussion exists in subsequent chapters. Therefore, the point of the following sections is to familiarize the reader with the current anti-spam techniques and proposals, in order to better understand the spam system that will be modeled in the next chapter.

### 3.2 Content Filtering

The most well known anti-spam solution category is content filtering. In this category, solution types examine and rate each message as to whether their content appears to resemble spam or not. If given a spam rating, then the message may be filtered in one way or another, which could range from immediate deletion to sequestration into another area or folder.

The most common content filtering type is rule-based content filtering. In this type, a set of rules or tests are run against the content of the message. Each test corresponds to some item in the message header or body that is associated with spam to some degree, e.g. exclamation points in the subject line. If a message meets the test, a value associated with the test is added to an overall spam score. After all the tests are run, a final spam score is achieved, which can then be used to filter messages as desired.

There are now rule-based content filters associated with most major email clients that are built-in to the clients. For example, all of Microsoft’s major email clients, Microsoft Outlook 2003, Hotmail, and MSN, all have intrinsic filtering systems (Microsoft, 2003). There are also many third-party filtering systems and providers as well (Open Directory, 2004). Of particular note is an open source filter called SpamAssassin, which is widely used by smaller ISPs and ESPs because of its open-source license (The Apache Software Foundation, 2004). SpamAssassin conducts a large array of tests on the message (The Apache Software Foundation, 2004). Often programs that use SpamAssassin will set a threshold score such that if messages get a score higher than the threshold, they are considered spam.

In addition to the static spam rules included with SpamAssassin, it also includes a Bayesian learning filter, which enables the rule-based technique to become dynamic (Wikipedia, 2004). Once the Bayesian filter is set up, SpamAssassin gives a probability of whether a message is spam is not. This Bayesian filter is then updated with each message received, and learns over time from those messages. For each message reported as spam by the user, the characteristics associated with that message become more “spammy” in the filter, i.e. increase the probability that future messages with similar characteristics will have a higher spam probability. And conversely, for each message not reported as spam by the user, the characteristics associated with that message become less “spammy” in the filter, i.e. decrease the probability that future messages with similar characteristics will have a higher spam probability.

Major ISPs also outsource additional rule-based content filtering on top of their in-house filters. The leading provider is Brightmail, Inc., which includes Microsoft, AT&T, Comcast,

21 SpamAssassin is a top-level project at the Apache Software Foundation, one of the most reliable sources of open source software and the developers of the Apache Web Server (Kerner, 2004).
Verizon, and Earthlink among its customers. Brightmail has come up with some innovative techniques to generate useful rules in addition to static spam tests and Bayesian filters. One such technique is the creation of “spam traps;” a spam trap is an email address set up for the sole purpose of collecting spam. Since no human exists behind these email addresses, it is assumed that all mail received at these addresses are spam. They also have honed the system of distributing new rules from their servers to email clients across the globe in almost real time (Brightmail, 2003).

A variation of rule-based content filtering is collaborative content filtering. In collaborative content filtering, messages are also filtered as to whether they are considered spam or not, but this consideration is made by actual people in collaboration as opposed to pre-defined rules. Some filtering software providers employ staff to mark spam messages as they come in, and set up dynamic rules to block future messages of the same type. More distributed approaches have been set up, however, where members of a spam filter community declare messages as spam, and this feedback instantly promulgates throughout the other users of the filter (Cloudmark, 2004).

Other anti-spam content filtering solution types consider specific pieces of email content as opposed to the content of the entire message. The client white list is perhaps the most familiar to the average reader. This technique is widely deployed and can be found in most major email clients. All it does is ensure that mail from particular email addresses does not get filtered and goes right through to main folders. A system where friends, family, and colleagues get filtered into individual folders would be a variation of the client white list. Another variation is to strictly apply the white list such that only email purported to be sent by those email addresses on the list gets delivered to regular inboxes—all other email would go to a junk folder.

Earthlink has taken the client white list concept one step further, and is the first major company to employ the “challenge-response” type on a large scale (Popov, 2004). In this type, any email that comes from someone not on a user’s white list is sent an automatic reply challenging them to identify themselves as someone that the user wants to receive email from. If the sender responds, this response can be verified by the receiver, and if accepted, the sender will be placed on the user’s white list. Earthlink has chosen not to send challenges to bulk mailers, however, and all this mail thus gets put in a suspect folder.

A final type in this category is client black lists or “block lists” that prevent certain senders from sending email to non-junk folders. For senders that repeatedly use the same address, this type serves as an easy alternative to asking the sender to stop sending email through an opt-out process. Since it easy to create such lists within email clients, most email clients have the ability to create client black lists.

Of all the general spam solution categories, content filtering has the most academic activity as well as commercial activity. Such academic work centers on the best ways to tune neural networks and Bayesian filters to detect spam as well as statistical work on how the spam problem is progressing. There is not yet a spam journal that is dedicated to this topic, but there have been several spam conferences where work has been presented.

---

22 Brightmail was bought by Symantec Corporation for an estimated $370 million in 2004 (Symantec, 2004).
For instance, at the Spam Conference 2004 at MIT, Terry Sullivan presented a statistical analysis on how spam changes over time, and Shlomo Hershkop introduced the idea of behavior based filtering where filtering could be done based on deviations from the user's normal email receiving habits (Houbart, 2004). Industry players like Microsoft and Brightmail, Inc., also gave talks about their latest content filtering practices.

3.3 Access Filtering

The second category of anti-spam solution types is access filtering. In this category, solution types attempt to block spam senders from accessing email servers, i.e. filtering access to only legitimate senders. As such, this category is the realm of Internet Service Providers (ISPs) and Email Service Providers (ESPs).

ISPs and ESPs face the spam problem on two fronts: (1) the more email users they have, the more a target they are for spam; and (2) the bigger they are, the easier it is to try to send spam from their systems because it is harder to monitor and control their systems. In 2003, some of the largest ISPs and ESPs got together to work collaboratively on a solution and formed a group called the Anti-Spam Technical Alliance (ASTA). This alliance includes America Online (AOL), British Telecom (BT), Comcast, EarthLink, Microsoft, and Yahoo!

On June 22, 2004, the first fruits of this collaboration appeared in the form of a document entitled “Anti-Spam Technical Alliance Technology and Policy Proposal” (ASTA, 2004). This document is intended to be a working document, and is versioned 1.0. It primarily enumerates lists of “Best Practices” for ISPs, Bulk E-mail Senders, and Consumers. The members have begun to implement these practices, with immediately encouraging results. For example, Comcast began blocking outbound email for some of their broadband customers' computers that they believe have been hijacked to send spam, and they have reduced outbound spam from their network by an estimated 35% as a result (Hu, 2004). This is particularly encouraging because recent estimates of spam sources indicate that up to 80% come from such hijacked machines, and Comcast is one of the largest sources of these machines (Leyden, 2004).

The best practices in general all attempt to reduce the possibility of spammers using third-party IP addresses to send their spam. Several practices specifically have to do with cutting down on hijacked machine use, either by literally shutting off consumer machines from sending bulk email in general, or by creating automated complaint systems to identify violators more quickly in order to shut them off individually as necessary. They also recommend better controlling of access to Web-based email accounts, such as Hotmail, so that spammers cannot create thousands of such accounts to send millions of emails that appear to come from the IP addresses of those services.

Other access filtering solution types include those that block access to bulk senders that do not meet certain criteria. One such proposed type is postage costs, which is a subset of the more general “sender pays” approach to spam where more of the costs of sending email are shifted to the sender. The idea of postage costs is to make postage work analogously for
email as it does for regular mail, where bulk mail senders have to pay postage that limits their sending somewhat. As such, it would require a significant overhaul of the email sending system to account for such charges (Rideau, 2004).

There has been much talk about postage costs, but no major proposals to speak of (Mara, 2004). However, there are small commercial vendors pushing the idea. For instance, Goodmail Systems, Inc. has been trying to sign up major e-mail providers such as Yahoo to charge bulk mailers a penny a message in return for a guarantee that their messages will pass spam filters (Goodmail Systems, 2004).

Very similar to the postage costs idea is that of computational costs, where bulk senders have to make their computers perform some computations, e.g. a complex math problem, before sending a message. Computational costs can be thought of as a type of postage, but I have distinguished them because postage is usually thought of as monetary. In fact, the most press postage costs has received in recent times stems from a comment by Bill Gates in early 2004 that was really about computational costs (Associated Press, 2004). A Microsoft Research Team has apparently been working on this idea since 2001.

The theory behind computational costs is that just as postage would limit bulk mail sending rates because of limited available funds by bulk mailers, so would computational costs because time is money. Less mail could be sent from one machine due to limited computational power, and thus more machines would need to be bought to achieve the same sending rate. Additionally, computational costs may be theoretically easier to implement because money does not need to be exchanged.

Perhaps the most widely deployed access filtering solutions, however, are those of server white lists and server black lists. These lists are analogous to the client white lists and client black lists discussed in the previous section, but these lists occur on the server. The bigger ESPs have created “white-lists” whereby legitimate bulk email senders, e.g. Amazon.com or Hallmark.com, can ensure their email is delivered. By itself, this would seem to make the spam problem worse or at least just the same, but these ESPs have also turned up their content filters for bulk senders. The result is that if you do not go through their white-listing programs, you are significantly at risk to having your email automatically go to junk folders.

Microsoft itself claims not to have an in-house white list, and instead out-sources their white list program to IronPort Systems’ Bonded Sender Program (IronPort Systems, 2004). This program requires bulk senders to put up a bond and have their email complaints monitored by IronPort, which deducts money from the bond for each complaint over a threshold. With Microsoft’s endorsement, this program has become more important, with it now covering an estimated 230 million email addresses (IronPort Systems, 2004). Many smaller ESPs that cannot afford to institute their own processes simply out-source this mechanism to IronPort. AOL (America Online, 2002) and Yahoo! (Yahoo, 2004) maintain their own white list processes.

As noted, there can also be black lists at the server level, either at the sending server, receiving server, or both. These lists try to, in real time, identify sources of spam such that ISPs and ESPs can refuse mail from them before any content filtering takes place. There are
many black lists, and they vary in their criteria for listing (Open Directory, 2004). Most are based upon individual IP addresses or ranges of IP addresses.

Two lists worth mentioning are the Open Relay Database (ORDB) (Open Relay Database, 2004) that tracks unprotected mail servers and the Realtime Blackhole List (RBL) (MAPS, 2004) that tracks known sources of spam. It is somewhat unclear which lists in particular are used by particular ISPs and ESPs, although it is known that Microsoft has used the RBL in the past for Hotmail (Festa, 1999). When an email server uses a black list, it usually prevents email from originating right at the beginning of the SMTP exchange, often by refusing connections altogether, or by sending back a negative message about the block.23

Another type of access filtering is a Do-Not-E-Mail Registry, which could work analogously to the Do-Not-Call Registry recently established in the United States (Federal Trade Commission, 2003). There are a number of different ways the registry could work in particular, but fundamentally it would make illegal sending unsolicited commercial email to a large list of email addresses. In other words, it would filter access to the email addresses listed in the registry.

In June, 2004, however, the FTC reported to Congress that they did not think the Do-Not-Email Registry would be a good idea at the present time (Federal Trade Commission, 2004). They concluded that because of the inability to reliably authenticate email senders and the penchant of spammers to ignore laws and regulations, an email registry would do more harm than good at this time. They noted that if the authentication situation were to change at some point in the future, the idea could be revisited.

A final type of access filtering widely deployed is the reverse DNS test. DNS stands for Domain Name System, and is the system whereby IP addresses are mapped to domain names (Wikipedia, 2004). That is, DNS is what enables one to type in www.mit.edu into a Web browser and consequently reach the MIT Web page. In this case, the Web browser uses DNS to look up the IP address associated with www.mit.edu, and then connects to the server behind that IP address to retrieve the Web page.

Reverse DNS is simply the reverse of DNS, and translates an IP address into a domain name. The way it is used in access filtering is in the initial HELO part of the SMTP exchange, as outlined in the last chapter. When the domain name is transmitted after the HELO command, the email server can use reverse DNS to find out what IP address this domain name should correspond to. If the IP address that was used to connect to the email server does not match the IP address found through reverse DNS, then the email can be considered suspect, and in the strictest of applications, the sender can be denied further access to the email server (see Figure 3).

23 Black lists can also be used by email clients in this manner, which look at the received headers, and then delete email upon receipt as identified. However, this method of implementing black lists in the client takes up more Internet bandwidth because the email actually was transmitted and received. Additionally, it requires the client to connect to a server to query against the latest list.
This method has been widely applied to high volume IP addresses or those that generate significant complaints. AOL uses reverse DNS in this manner (America Online, 2004). However, they stop short of strict application for all IP addresses because of a significant problem with doing so, which will now be explored. Recall that DNS translates a domain name, e.g. www.mit.edu, into an IP address, e.g.:

```
bash-2.05b$ nslookup www.mit.edu ns1.mit.edu
Server: W20NS.MIT.EDU
Address: 18.70.0.160
Name: DANDELION-PATCH.mit.edu
Address: 18.181.0.31
Aliases: www.mit.edu
```

In this example, the UNIX ‘nslookup’ command was used to ask the mit DNS server (ns1.mit.edu) what the IP address is for www.mit.edu, and it returned ‘18.181.0.31’. Reverse DNS asks the reverse question of what IP address is associated with a particular domain name.

```
bash-2.05b$ nslookup 18.181.0.31 ns1.mit.edu
Server: W20NS.MIT.EDU
Address: 18.70.0.160
Name: DANDELION-PATCH.mit.edu
Address: 18.181.0.31
```

In this second example, the same command was used to ask the same DNS server what domain name is associated with the IP address ‘18.181.0.31’, and it returned ‘DANDELION-PATCH.MIT.EDU’. This answer is unexpected since the original domain name queried was ‘www.mit.edu’. The reason for the discrepancy is that multiple domain names can point to one IP address in DNS, but only one domain name can be returned in the reverse DNS lookup.

In other words, the device connected to the IP address ‘18.181.0.31’ has several domain names associated with it, including ‘www.mit.edu’ and ‘DANDELION-PATCH.MIT.EDU’, the latter of which is the primary domain name according to DNS. There is simply not a
one to one mapping between IP addresses and domain names in all contexts. In situations
where this is not the case, not all email senders from those computers can meet the standard
of specifying the HELO line in such a way that it passes the reverse DNS test, and this is
why it is not strictly applied in all cases.

3.4 Identifiability

The third category of anti-spam solution types is that of identifiability. In this category,
solution types attempt to more reliably identify the senders of email messages in general.
The idea is that if senders can be reliably identified, then:
1. Spam senders can be more easily traced and thus dealt with; and
2. Content and access filtering can be improved because they can be based on more
   reliable sender information.

The first solution type in this category is that of IP address authentication. Recall from the
previous chapter that the IP address is currently the only piece of information provided by
the sender during the SMTP exchange that is somewhat trustworthy. IP address
authentication attempts to use that fact to restore the trust that was once shared by email
users that email received is really from who it says it is from.

There are many major proposals that are competing for implementation in this area. The
proposals have been submitted to the Internet Engineering Task Force (IETF), which has a
working group to consider all the proposals and make some recommendations (Newton et
al., 2004). The two leading proposals merged into a proposal called Sender ID (Microsoft,
2004). The proposals that merged were Sender Policy Framework (SPF) by Meng Wong
from pobox.com and Caller ID by Microsoft (IC Group, 2004). AOL and many other
domains have become early adopters of this system (America Online, 2004).

All of these proposals basically work the same way, which will now be briefly explained.
They use DNS and the reverse DNS concept of publishing some information in DNS that
the email server or client can then use to authenticate the sender via their IP address. The
main difference between the proposals and the regular reverse DNS test is that the proposals
authenticate parts of the message, while the reverse DNS test authenticates just the IP address
that is connecting to the email server.

For example, the from (sender's) email address can be used. In this case, the domain part of
the address is first extracted. The sender would have previously published in their public
DNS record which IP addresses are authorized to send email for the domain extracted.
Those authorized IP addresses are then looked up through DNS. If there is a match, then
the message passes; otherwise, the message can be considered suspect (see Figure 4).
IP Address Authentication

1. IP Address A
2. FROM x@example.com
3. DNS
4. example.com
5. If IPA in IP Set, Then continue, Otherwise Maybe do Something else

Sender's Email Client

Sender's Email Server

A similar type to IP address authentication is that of content signing. The leading content signing proposal has been put forth by Yahoo! and is entitled DomainKeys (Yahoo, 2004). Instead of publishing in DNS the IP addresses a domain is allowed to send from, DomainKeys has one publish a public encryption key, which is part of a public/private key pair that can be used to encrypt and decrypt messages (Wikipedia, 2004).

When messages are sent from a domain, the sender or their email server uses the private key associated with that domain to create a digital signature of the message, and stores this digital signature in the header of the message. Then the email server that receives the message attempts to verify this signature using the public key in DNS associated with the domain of the 'from' address of the message. Again, if there is a match then the message passes, and otherwise it can be considered suspect (see Figure 5).

Content Signing

1. Sign Message
2. FROM x@example.com
3. DNS
4. example.com
5. If signature valid, Then continue, Otherwise Maybe do Something else

Yahoo! has also submitted their proposal to the IETF, and is working with Sendmail, Inc., the maker of the most popular email server, to develop a reference implementation (Delany, 2004). Eric Allman, the CTO of Sendmail, Inc., has indicated that Sendmail will support all authentication schemes that gain traction (Lang, 2004).
Content signing is thus similar in basic structure and result to IP Address Authentication—they both attempt to authenticate emails by comparing something given by the sender with specific parts of messages received, and they both are after this same result of increased trust. However, they are not mutually exclusive such that both could be used against a particular message in conjunction.

Both Sender ID and DomainKeys also resolve the aforementioned problem with the traditional reverse DNS check. They do this by avoiding the traditional reverse DNS call, and instead use additional information in regular DNS for authentication. Consider the problem case before of multiple domains behind one IP address. With these new proposals, each of those domains can publish their own IP addresses and/or public keys in DNS, and some program wishing to authenticate them can find the information associated with the particular domain in question.

3.5 Regulation

The fourth category of anti-spam solution types is that of regulation. This category includes public laws as well as private contracts that prohibit spam techniques. Many US states have laws on the books that apply to spam in one way or another. General laws include consumer protection and computer fraud codes, which sending spam often violates. Nevada was the first state to enact a spam law in particular (in 1997), and many other states followed in their footsteps (Sorkin, 2004). Most of these laws contain similar provisions, which are summarized below.

- They may outlaw falsified routing information, such as faking the sender email address or "spoofing" the IP address.
- They may require that unsolicited commercial email be labeled with a particular tag, e.g. "ADV:", in the subject line.
- They may require that unsolicited sexually explicit email be labeled with a particular tag, e.g. "ADV:ADULT", in the subject line.
- They may prohibit misleading or fraudulent message subjects or bodies.
- They may require "opt-out" instructions that enable Consumers to stop receiving messages from a particular mailing list or sender.
- They may require a valid "reply-to" address, postal address, or phone number so that Consumers can get in touch with the sender if desired.

However, due to the cost of identifying spammers and jurisdictional problems, not many cases have been brought against spammers under these laws. The most notable states that have taken quite active positions are the Washington State (Washington Office of the Attorney General, 2004) and Virginia (Virginia Office of the Attorney General, 2004). Both states enforce their spam laws via their Attorneys General's offices.

One notable case resulted in the first significant jail sentence for a spammer. The New York State Attorney General Elliot Spitzer convicted a spammer under the state's identity theft law, resulting in 3.5 to 7 years of imprisonment, which is the maximum allowed under the law given the defendant's prior record (Roberts, 2004).
The CAN-SPAM Act of 2003 pre-empted state spam laws (except their anti-fraud provisions), and in so doing, became the primary law in the United States in terms of directly regulating spam. This law incorporated much of the common provisions in the state laws, including the following.

- Requiring unsolicited commercial email to have opt-out methods.
- Requiring unsolicited commercial email to include the sender's physical address.
- Prohibiting deceptive subject lines.
- Prohibiting false headers.
- Requiring labeling of sexually explicit email (Federal Trade Commission, 2004).

The Act does not go so far as to prohibit unsolicited commercial email in general, nor does it apply to non-commercial spam. Furthermore, it is to be expected that the law will not make much of a difference until it is significantly enforced. Enforcement, as stipulated by the Act, is directed by the FTC or by willing ISPs. The FTC filed its first suit under the law in April 2004 (Federal Trade Commission, 2004). The law actually took effect on January 1, 2004. ISPs, namely AOL and Microsoft have also filed several suits and are expected to be among the main enforcers of the law.

The proposed sentencing guidelines associated with the law are rather steep, and include imprisonment for up to five (5) years for some violations (United States Sentencing Commission, 2004, pg. 155). While no one has yet faced this type of punishment under this law, the FBI has told Congress it has identified about 100 significant targets, and is actively working to bring them to justice (McCullagh, 2004).

Furthermore, many countries besides the United States have enacted their own spam laws (Sorkin, 2004). The particular provisions vary from country to country. One notable difference that has been talked about considerably in the press is that the EU laws require commercial senders to get recipients to opt-in before sending, while the US law does not (message labs.com, 2004). Since spam is an international problem, these differences cause considerable trouble for compliance by global companies, but are largely ignored by more lawless spammers.

In addition to public laws, each ISP and ESP has developed their own anti-spam regulations to help combat spam, which are usually written into their Acceptable Use Policies (AUP). ISP and ESP AUPs, for the most part, all prohibit spam, giving them the ability to terminate accounts if they feel customers are sending significant amounts of spam. In conjunction with AUPs, these organizations generally have set up complaint systems whereby their email customers can report spam. These complaints get sent to some centralized location, where they can be aggregated in order to try to identify the most egregious perpetrators.

### 3.6 International Collaboration

International collaboration is the fifth category of anti-spam solution types. The result of such collaboration would be some type of regulation, which is the fourth category. I have included international collaboration as its own category because it cuts down on the ability
for spam senders to escape regulation by moving jurisdictions. There has not yet been much international collaboration to speak of, however.

The UN has decided to get involved in this issue, but at a rather slow pace. The General Assembly adopted resolution 56/183 in 2001 that called for a world summit on the information society to be held in two phases (United Nations, 2002). The first phase was held (December 2003) in Geneva, Switzerland and resulted in a Declaration of Principles and a Plan of Action (World Summit on the Information Society, 2004). In the Plan of Action document, spam received this mention: “[t]ake appropriate action on spam at national and international levels.” While this is not very specific, it does indicate that there exists support to work on this problem via international consensus.

3.7 Enforcement Capacity

The final and sixth listed category of anti-spam solution types is that of enforcement capacity. In this category, solution types attempt to increase the enforcement of existing regulations. The most obvious way this is done is through lawsuits. States taking spammers to court under criminal codes has already been mentioned. In addition to those criminal lawsuits, Microsoft and the bigger ISPs and ESPs have also been taking their spam problems to court via civil suits.

On March 10, 2004, Microsoft, AOL, Earthlink, and Yahoo! jointly announced legal suits in federal courts in California, Georgia, Virginia, and Washington State. These suits were filed under the Controlling the Assault of Non-Solicited Pornography and Marketing (CAN-SPAM) Act of 2003, which as noted, took effect January 1, 2004 (United States of America, 2003). These were over 200 separate cases, although most of them were filed against “John Does” because the identities of the alleged perpetrators are unknown. More cases of a similar nature have followed since then and are expected to continue.

These suits and the ones mentioned in previous sections are not the first spam suits to be filed, however. AOL has been perhaps the most active legally, and has been filing suits since at least 1996 (America Online, 2003). Some of these suits are under state spam laws, and others are under consumer protection laws and in some cases tort common law.

One notable recent case in this latter category was a case brought by Intel alleging a former employee committed a tortuous trespass when he sent bulk email to Intel employees. The case made its way up to the California Supreme Court, which decided against Intel claiming that Intel did not show real property damage required for liability under the trespass to chattels cause of action sought (Boulton, 2003).

Sometimes, however, suits are foregone by vigilantes attempting to take the law into their own hands and shutting down spam senders through private enforcement. Vigilante efforts are really a catch-all for the attempts by spam receivers to stop receiving spam beyond the more traditional techniques they have at their disposal. These efforts include:

- Attempting to identify the true senders of spam.
- Reporting spam messages to ISPs, ESPs, and governmental authorities.
- Spamming back the people they believe are spamming them.
- Posting spam received to public newsgroups and Web sites.
- Attempting to publicly shame senders of spam.
- Creating customer service nightmares for senders of spam.
- Denial of Service attacks of Web sites and servers associated with spam sending.

In general, these efforts are designed to increase the "cost" of sending spam by adding some more friction to the system. However, vigilante efforts can easily escalate across the line into illegality (Hollander, 2004) as well as cause trouble for innocent bystanders (Fausett, 2004). For instance, vigilantes have been known to break laws, even spam laws, in the process of trying to stop spam senders. Also, less sophisticated vigilantes often falsely attack innocent third-party victims of identity theft by spam senders. However, when carefully controlled, these efforts can be quite useful. The most notable example of useful collaborative vigilante effort is the Spamhaus Registry Of Known Spam Operations (ROSKO), which compiles evidence of spam operations for use in court or otherwise (Spamhaus, 2004).

A final type of solution that attempts to increase enforcement capacity is that of a bounty system. The CAN-SPAM Act of 2003 instructs the FTC to consider such a system (Brunker, 2004). As Brunker explains in an MSNBC article:

> As outlined in CAN-Spam, the bounty system would offer a person who first identifies someone violating the law's provisions a reward of "not less than 20 percent of the total civil penalty collected" by the FTC.

> The FTC is compiling and reviewing expert testimony on the bounty plan and will report back to Congress by September on whether the idea is viable. The study was mandated by a little-noticed 11th-hour addition to the law by Sen. Jon Corzine, D-N.J., and Rep. Zoe Lofgren, D-Calif.

The system was originally proposed by Stanford Professor Lawrence Lessig, and would essentially legitimize most vigilante efforts, and would further incentive these efforts with monetary payouts. The policy would effectively try to reroute consumer technical expertise (and time) to help the FTC identify spammers. Not surprisingly, email marketers are not too fond of the idea (Fadner, 2004).

### 3.8 Spam Solution Activity Summarized

As you can see from the above sections, there are a myriad of approaches to solving the spam problem, including technical, legal and policy oriented approaches. A summary can be found in the first section (see Table 4).

When reading through these different approaches, one can become overwhelmed with the depth and breadth of activity taking place. However, I will show that this solution activity can be rather easily placed within a relatively simple model of the spam system.

Many of the solutions are similar in the pressures they put on this system, and that is why they have been grouped into a smaller number of spam solution categories in this chapter.
These categories prove useful when thinking about an effective spam solution strategy. In the next chapter, the model of the Base Spam System will be introduced.
In this chapter, a model of the Base Spam System is presented, which is a pictorial model of the causes and effects that comprise the intrinsic dynamics of the spam problem (see Figure 6). This model will be used in the next chapter to more thoroughly evaluate the solution activity presented in the last chapter.

This model of the Base Spam System is intended to be the central contribution of this thesis. It stems from recognition that the spam problem is really a complex system, and it is an attempt to capture the complexity of that system. The major insight is that, once the system is depicted, one can begin to think about solutions to the spam problem in a holistic manner. That is, one can use the Base Spam System model to evaluate how specific solution activity might affect the spam problem, which is done in the next chapter.

The primary desired properties of the model are:

- To be able to visually communicate the complexities of the spam problem.
- To be able to expand the model to include complexities that may arise in the future.
- To be able to evaluate spam solution activity.
- To be able to analyze the spam problem both qualitatively and quantitatively as needs arise.
These properties led to a choice to model the Base Spam System in the form of a causal loop diagram, which is a type of diagram used in the discipline of system dynamics. That methodology will now be explained, and then the model will be presented thereafter.

4.1 Causal Loop Methodology

The causal loop diagram is a methodology used in the discipline of system dynamics. It is a diagram of a system comprised of the variables of that system linked together to depict the causal loops and linkages in the system. Perhaps the best introductory explanation of the causal loop methodology is from John Sterman's book *Business Dynamics*:

A causal diagram consists of variables connected by arrows denoting the causal influences among the variables. The important feedback loops are also identified in the diagram...

Variables are related by causal links, shown by arrows. ... Each causal link is assigned a polarity, either positive (+) or negative (-) to indicate how the dependent variable changes when the independent variable changes....

A positive link means that if the cause increases, the effect increases above what it would otherwise have been, and if the cause decreases, the effect decreases below what it would otherwise have been...

A negative link means that if the cause increases, the effect decreases below what it would otherwise have been, and if the cause decreases, the effect increases above what it would otherwise have been...

Link polarities describe the structure of the system. They do not describe the behavior of the variables. That is, they describe what would happen IF there were a change. They do not describe what actually happens...

...When assessing the actual behavior of a system, all variables interact simultaneously (all else not equal) and computer simulation is usually needed to trace out the behavior of the system and determine which loops are dominant. (Sterman, 2000, p. 138-140, emphasis kept from the original text)

The causal loop methodology closely aligns with the desired properties of the model stated in the last section. Casual loop diagrams are designed to communicate complexity visually. And because of their pictorial nature, they are easily expanded or changed to adapt to changing realities or to include additional behavior such as the side effects of spam solution activity. The diagram itself is qualitative in nature; however, the linkages between the variables imply quantitative relationships. These relationships can be studied in future work.

There are two important things to keep in mind when thinking about causal loop diagrams in general and the model of the Base Spam System in particular. First, the model is not verified to be complete in a theoretical sense. Instead, it is complete in an empirical sense in that it has captured all of the causes and effects observed to date. To the extent that more causes and effects are observed, the model can be easily expanded to include them.

Second, the variables in the model are a choice. The model depicts causal linkages and loops, and those underlying causal chains can be depicted in a number of ways, each using a different set of variables depending on how detailed one wants to be with regards to intermediate causal steps and which of those steps one wants to highlight. In this model, the variables have been chosen in order to best simply communicate the complexities of the
spam problem. That is, the particular variables (and their names) have been selected to create a model that is as easy to understand as possible.

I believe that the best way to grasp the causal loop methodology is through example. This chapter provides many examples, building up the model that is the Base Spam System in steps, starting with the central Spam Ratio variable in the next section.

### 4.2 Spam Ratio

Recall that the spam ratio is the metric introduced in Chapter 1 to evaluate the spam problem for a particular user of email. This ratio is the amount of spam received divided by the total amount of email received. The central variable of the model of the Base Spam System is the Spam Ratio in the aggregate, i.e. the total amount of spam received across the Internet divided by the total amount of email received across the Internet (see Figure 7).

**Figure 7 Spam Ratio**

Two variables directly influence the Spam Ratio, which you can see in the model by the two arrows pointing to it. These variables are Spam Sent and Non-Spam Sent. The plusses and minuses explain the direction, or polarity, of the influences. For instance, as Non-Spam Sent increases, all other things constant, the Spam Ratio decreases. Similarly, as Spam Sent increases, all other things constant, the Spam Ratio increases. Since I will be explaining
many linkages like these, please assume that when I am talking about the relationship between two variables, all other variables are held constant.

There are two other effects between the Spam Ratio and Non-Spam Sent that are also depicted. First, as the Spam Ratio increases, there is more and more spam in recipient inboxes. This means that more time must be spent dealing with this spam, which decreases the Desire to Use Email in general. Another way to look at this phenomenon is that if the time spent on email remains constant, and there is more spam, less time can be used for sending Non-Spam, which is the desired use of email.

In any case, an increase in the Spam Ratio decreases the Desire to Use Email variable, which decreases Non-Spam Sent, and in turn, increases the Spam Ratio further. In the figure, this loop is identified with the title Loss of Usability. Next to the loop label is a curved arrow, describing the direction in which the loop occurs. The plus inside the curved arrow indicates that the loop is a reinforcing loop. That is, the result of the causal loop is a reinforcement of the original change, i.e. an increase or decrease in the Spam Ratio.

The second loop in the figure, identified with the label Loss of Trust, has the same net effect as the first loop but for different reasons. Recall that most spam employs fraudulent tactics in an attempt to hide the identity of the sender. These tactics often involve using other peoples' identities. In the aggregate, this leads to a general loss in trust that a given email is who it says it is from. From the perspective of non-spam senders, this leads to a decrease in the expectation that their email will be perceived correctly, noted by the Expectation Email Perceived Correctly variable. In turn, this leads to less non-spam being sent because legitimate senders have less trust that their non-spam will be perceived by recipients in a useful manner.

Therefore, as the Spam Ratio increases, the Expectation Email Perceived Correctly variable decreases, which decreases the Non-Spam Sent variable, and in turn, further increases the Spam Ratio. As such, this loop is also a reinforcing loop. Perhaps the best example of this loop is firms that have been major victims of spam identity theft scams. These firms may be more reluctant to communicate with their customers via email, in fear of simply exacerbating their email problems. Instead, they may opt to have their customers log in to their accounts, and then distribute news and information through their Web sites.

Finally, note that the denominator and numerator of the Spam Ratio involve email received and not email sent. In the model, however, the variables that impact the Spam Ratio are Spam Sent and Non-Spam Sent. These variables equal Spam Received and Non-Spam Received when there is no filtering taking place, and are greater than Spam Received and Non-Spam Received when filtering exists.

Yet the influences are the same. Consider that some percentage of all email sent is received, depending on the percentage of email filtered. Therefore, as Spam Sent increases, so does Spam Received (the percentage that gets through), and in turn, the Spam Ratio increases. In other words, it does not matter in terms of correctness whether Spam Sent or Spam Received is used as the variable.
However, I believe the choice of Spam Sent and Non-Spam Sent is better for two reasons. First, it allows for filtering to be considered more intricately, as will be done in the next chapter. Also, Spam Sent and Non-Spam Sent are clearer because they more basically correspond to actions taken in the real world, and therefore are easier to understand.

4.3 Spam Sent

As mentioned, one of the main variables influencing the Spam Ratio is Spam Sent. Two variables directly influence Spam Sent: Spam Capacity and Desire to Spam. Both Spam Capacity and Desire to Spam independently influence Spam Sent, and are thus depicted independently in the model. One way to see why the distinction matters is to consider that many people have the capacity to spam, but no desire to spam, and those people do not increase Spam Sent. Similarly, there are those that have desire to spam, but no spam capacity, and those people do not increase Spam Sent. But if someone has a slight capacity, and then his/her desire to spam increases dramatically, we can expect that Spam Sent will increase since he/she will spend more time to use his/her current capacity to increase Spam Sent given his/her strong desire.

This section considers the drivers of Spam Sent. First, consider a primary reinforcing loop of the Spam Sent variable via Spam Capacity, namely Ability to Spam (see Figure 8).
Spam Capacity is directly affected by Access to Methods such that as Access to Methods increases, Spam Capacity does so in turn. The methods contained within the Access to Methods variable include any method used by spam senders, including all of the techniques outlined in Table 3 (from Chapter 2) used to hide identity. Fundamentally, access to these methods requires money. This does not necessarily mean that spam senders need to pay others for access; they could just need money to free up their time so that they can learn to access the methods themselves. In any case, Access to Methods can increase as a result of Profits increasing. Note that it is not just money that drives Access to Methods, but Profits, because without profiting from Spam Sent, spam senders would rather spend their money on more profitable activities.

Continuing to follow the loop around backwards, we see that as Responses increases, so does Profits in turn. Responses is an absolute measure of the responses that people make from Spam Sent to spam senders that generate Profits. Finally, as Spam Sent increases, so do Responses, because as there is more and more Spam, there will be me more and more Responses, assuming a non-zero response rate and all other variables held constant.

This loop, in and of itself, is reinforcing because an increase in Spam Sent would serve to increase Spam Sent indefinitely absent other effects. This may seem like it is the case as Spam Sent has increased steadily, but we know that Spam cannot increase indefinitely. As such, there is a balancing loop, labeled Saturation in the model.

The Saturation loop includes the Response Rate variable, which is simply the percentage (%) rate of response from Spam Sent. As the Response Rate increases, so does Responses. Yet as Spam Sent increases, the Response Rate actually decreases because those responding have simply more offers to consider. This saturation behavior is a familiar economic concept seen in traditional advertising and other markets.

Next, let us concentrate on Profits and expand our thinking on this variable (see Figure 9).
The Innovation loop depicts how an increase in Profits can indirectly increase Access to Methods in addition to direct ways such as freeing up time and paying others already mentioned. In this loop, some Profits are allocated towards Research and Development (R&D) such that as Profits increase, all other things held constant, so does R&D. In turn, R&D results in wholly new methods that spammers then have access to. This loop is similar to R&D present in manufacturing companies where some portion of their profits is expended to improve on old methods and come up with new manufacturing methods.

Profits also influence the Desire to Spam variable with a positive polarity. Note that there are other variables that may increase or decrease Desire to Spam, but all other things held constant, an increase in Profits from spamming will lead to increase in Desire to Spam as more of those Profits from spamming are sought. Such an increase in desire leads to an increase in Spam Sent both directly and indirectly. The direct link is noted by the loop labeled Motivation in the model, capturing the act of an increase in motivation to use existing methods more in response to increasing desire.

The indirect link is noted by the loop labeled Initiative in the model, capturing the act of taking initiative in response to increasing desire. As the Desire to Spam variable increases, so does the R&D variable in turn, as the increased Desire to Spam translates into action. That is, not only does an increase in Desire to Spam directly increase Spam Sent, but it also indirectly increases Spam Sent via Spam Capacity as some of that increase in desire is funneled into coming up with new methods to spam.
4.4 Social Reaction

Thus far, we have primarily considered loops that work to reinforce the central Spam Ratio variable. We now turn to the major balancing loops that serve to balance these reinforcing loops. This section considers the loop identified as Social Reaction (see Figure 10).

Figure 10 Social Reaction

As the Spam Ratio increases, so does Recipient Concern for the spam problem. Increased Recipient Concern brings media coverage and a general negative feeling towards those that send spam. From the perspective of the spam sender, an increase in Recipient Concern, all other things constant, increases Spammer Alienation. Those who send spam are thus less likely to want to talk about what they do, and are more likely to engage in further activity to hide their identities and spamming from those around them. Few people desire to live an alienated lifestyle, and thus, all other things constant, and increase in Spammer Alienation leads to a decrease in Desire to Spam. As Desire to Spam decreases, as we have seen, Spam Sent decreases, and in turn, the Spam Ratio decreases, balancing the original supposed increase.

24 The only balancing loop we have considered is the Saturation loop.
The other major balancing loops relate to the variable of Regulation. As Recipient Concern increases, Regulation increases in response to this concern. Regulation includes not only public laws, but also private contracts that are related to spam sending, such as the Acceptable Use Policies (AUPs) agreed to by customers of Internet Service Providers (ISPs) and Email Service Providers (ESPs).

First, consider the balancing loop labeled Regulative Reaction (see Figure 11). As Regulation increases, Enforcement Capacity increases as well. Enforcement Capacity includes not only authorization to enforce from Regulation, but also the capacity to enforce in terms of resources needed such as manpower, knowledge, funds, etc.

As Enforcement Capacity increases, Enforcement Success increases, and as Enforcement Success increases, Perceived Consequences increases. The Perceived Consequences variable is related to the fines, jail time, and black listing that may result from lawsuits and private enforcement. In other words, as Enforcement Success increases, so does Perceived Consequences, which in turn, decreases Desire to Spam. As we saw in the last section, a decrease in Desire to Spam results in a decrease in Spam Sent that leads to a decrease in the Spam Ratio, which balances the original supposed increase in Spam Ratio that fueled the Regulation.
The two other loops in the figure reinforce and balance the Regulative Reaction loop. The loop labeled International Problem balances an increase in Perceived Consequences due to an increase in Enforcement Success. To see this loop in practice, consider Enforcement Success of the US federal spam law increasing dramatically, which would increase Perceived Consequences for those operating in the jurisdiction covered by the law. One way for spammers to lessen this effect on Perceived Consequences is to move overseas in an attempt to escape US prosecution. Such movement is captured in the Move to Less Regulated Areas variable, which when increased, leads to a decrease in Enforcement Capacity.

The loop labeled Precedent reinforces Enforcement Capacity as a result of an increase in Enforcement Success. Discrete increases in Enforcement Success are particular private or public cases against spammers that result in some consequences for those spam senders. Each successful case makes the next analogous case easier to pursue and win. That is, successful cases define precedents that increase Enforcement Capacity.

The other highlighted variable in the figure is that of Legal Protections, which balances Regulation. While Regulation can and is created as a result of Recipient Concern, such Regulation has limits, namely those created by Legal Protections. In the US, there are constitutional protections that limit the extent of federal and other laws. For example, free speech protection written into the First Amendment to the US Constitution has a direct effect on what kind of spam legislation can be created. State and other countries’ constitutions can have similar effects.

Finally, consider the loop labeled Regulative Restriction (see Figure 12). This is the other major balancing loop that results from the Regulation variable. Regulation, by definition, restricts behavior, and in this case, such restriction is related to Access to Methods. An increase in Regulation thus results in a decrease in Access to Methods, which decreases Spam Capacity. As we have seen, a decrease in Spam Capacity decreases Spam Sent, which balances the Spam Ratio increase that initiated the original increase in Regulation.
4.6 **Base Spam System**

At this point, we have seen all the different pieces of the Base Spam System, which are the variables connected together by cause and effect that comprise the spam problem. By "base," I mean the intrinsic dynamics of the spam problem. Since many loops and variables have been discussed so far in this chapter, I think it is important to engage in a short recap given the complete model of the Base Spam System (see Figure 13). At the center is the central Spam Ratio variable, with Non-Spam Sent and Spam Sent connected to it in the described manner. Then affecting Spam Sent, is Spam Capacity and Desire to Spam, both with positive polarity.
The loops identified as Ability to Spam and Motivation drive Spam Capacity and Desire to Spam, respectively. Additionally, they are both balanced indirectly by the loop identified as Saturation, which influences Responses, and in turn, Profits. Profits also directly drive the Innovation loop, and indirectly the Initiative loop via Desire to Spam.

And then there is the major balancing loop of Social Reaction related to Spammer Alienation. Finally, there are the loops discussed in the last section related to Regulation. The Regulative Restriction loop balances the Spam Ratio via Access to Methods, and the Regulative Reaction loop does so via Perceived Consequences. The International Problem and Precedent loops further influence this Regulative Reaction loop.

As mentioned, what can be added to this model of the Base Spam System are the effects from the spam solution activity outlined in the last chapter. Those effects will be the subject of the following chapter. Before we discuss that activity, however, we can step back and ask a more fundamental question, namely: what are the levers that can be pulled in the Base Spam System that would influence the Spam Ratio positively or negatively, either indirectly or directly?
Consider the variables in the model of the Base Spam System that can be altered by external forces (see Figure 14). These variables are effectively levers that can be pulled by spam solutions or otherwise to influence the system and thus the spam problem.

When one first examines the model for possible levers, one might think that every variable is a lever. For example, one might imagine decreasing Spam Capacity or Desire to Spam directly, or for that matter, decreasing Spam Sent directly, which straightforwardly influences the Spam Ratio. But upon further examination, one sees that those variables are not directly movable from the outside, and instead must be dealt with indirectly.

Consider Spam Sent, which is a variable fundamentally driven by spammers out in the real world sending spam. Unless one can physically stop spam senders from typing on their computers, then one can not really directly influence Spam Sent from the outside. Instead, what one can do is influence another variable that influences Spam Sent, such as Access to Methods.

Similarly, one cannot realistically influence Desire to Spam directly from the outside, which would involve physically changing the emotion of desire within the mind of spammers. Instead, one can raise awareness of the problem by increasing Recipient Concern, which in turn, would increase Spammer Alienation, and finally decrease Desire to Spam as wanted.
Likewise, instead of influencing Spam Capacity directly, one can decrease spammers' Access to Methods, which would in turn, influence their Spam Capacity in the desired direction.

Recipient Concern and Access to Methods are indeed two of the levers identified, Levers H and B, respectively. A third lever, and perhaps the most obvious point of external influence, is to decrease the Spam Ratio directly by attempting to remove spam before it reaches the inbox, but after it is sent. This is the arena of content filtering and occurs as a result of pulling Lever A.

A fourth lever, Lever C, involves influencing the Response Rate variable. Doing so is similar to influencing Recipient Concern, and in fact efforts to do one may spill over into the other. For example, decreasing Response Rate involves convincing those who are responding to spam to stop responding to spam, or at least decrease their response rate in the aggregate. A similar lever is Lever D, which is also about educating people. In this case, pulling Lever D involves educating spam senders about the consequences of their actions, which in effect, influences the Perceived Consequences variable from the outside.

The other levers (E, F, and G) involve the left side of the figure and the regulative process. One might be inclined to say one can increase Enforcement Success directly, but this is not realistic. Enforcement Success depends on the workings of established processes, and cannot be easily influenced directly. Instead, we can think about enacting new Regulation (Lever G), which directly increases Enforcement Capacity, which in turn eventually leads to the desired effects on Enforcement Success.

One can also imagine influencing Enforcement Capacity directly (Lever F) through helping enforcement efforts by allocating more resources towards enforcement, for example. Finally, we can also attempt to counteract the International Problem directly by attempts to directly decrease the Move to Less Regulated Areas variable (Lever E), through international treaty, for example. In the next chapter, we consider pulling these levers in a useful manner with regards to the spam problem.

---

25 One way Enforcement Success might be directly influenced, for example, is in the bribing of judges.
5 Spam Solution Interactions with the Base Spam System

In the previous chapter, I presented a model of the Base Spam System. With that tool at our disposal, we can begin to make sense of the major spam solution activities presented in Chapter 3 in terms of how they interact with the Base Spam System model. I will first reintroduce the spam solution categories. Then I will examine in more depth the interactions of each category with the model. In so doing, this fifth chapter uses the model to qualitatively evaluate spam solution types.

5.1 Spam Solution Activity Revisited

The following table (Table 5) re-lists the major spam solution types as presented in Chapter 3 (see Table 4) grouped by their categories as presented in that chapter. What is added is the second column, which reinforces why the category lines were drawn this way. One can see now that the reason for grouping the solution types in this manner is because the types within a given category interact with the Base Spam System similarly. In particular, each type within the same category pulls the same levers. I will now present each category in detail.

<table>
<thead>
<tr>
<th>Category</th>
<th>Lever(s) Pulled</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content Filtering</td>
<td>Lever A: Spam Ratio</td>
<td>Rule-based Filtering Collaborative Filtering</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Client White Lists Challenge / Response</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Client Black Lists</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Server White Lists Server Black Lists Do Not Email Registry Reverse DNS</td>
</tr>
<tr>
<td>Identifiability</td>
<td>Lever B: Access to Methods</td>
<td>IP Address Authentication Content Signing</td>
</tr>
<tr>
<td>International Collaboration</td>
<td>Lever G: Regulation</td>
<td>International Spam Protection Laws</td>
</tr>
<tr>
<td>Enforcement Capacity</td>
<td>Lever F: Enforcement Capacity</td>
<td>Lawsuits Bounty Vigilante Efforts</td>
</tr>
</tbody>
</table>
5.2 Content Filtering

As noted in Chapter 3, content filtering solutions attempt to declare messages as spam or not based on their content, and then filter them appropriately. With regards to the model, this type of action pulls Lever A, i.e. is an attempt to influence the Spam Ratio directly. In these cases, the alleged spam has already been sent. Then when the email arrives, alleged spam is filtered, reducing the Spam Ratio. Let me just briefly review how each of the solutions grouped into this category go about this task.

- **Rule-based Filtering** uses rules to declare what is spam or not via software.
- **Collaborative Filtering** relies on community processes in conjunction with collaborative software to declare what is spam or not.
- **Client White Lists** look at email headers, and decrease the probability that non-spam gets filtered.
- **Challenge / Response** sends challenges to all senders and then filters all email that do not have valid responses associated with them.
- **Client Black Lists** look at email headers to filter known bad senders.

When these methods work successfully, a reduction in the Spam Ratio is achieved (see Figure 15).

Figure 15  Content Filtering Effects

When the Spam Ratio decreases, the model predicts such a change will directly influence three other variables. First, we can expect both the Expectation Email Perceived Correctly...
and the Desire to Use Email variables to increase, which result in increasing Non-Spam Sent and eventually in another decrease in the Spam Ratio. That is, these loops, being reinforcing loops, reinforce our change and magnify it for the better. At the same time, as the Spam Ratio is reduced, Recipient Concern for the spam problem is also reduced. As this happens, we also expect less Regulation and Spammer Alienation to result, which could eventually lead to more Spam Sent, as the Social Reaction, Regulative Reaction, and Regulative Restriction loops take effect.

Whether the Spam Ratio actually is reduced after the system reacts to our initial change in this variable is unclear, and depends on the relative ratios between the effects mentioned. What complicates this picture even further is that pulling Lever A leads to side effects not present in the Base Spam System. To bear these out, let us add to this discussion all the possible interactions with the Spam System that can result from influencing the Spam Ratio directly (see Figure 16).

**Figure 16 Content Filtering Side Effects**

First, consider the loop labeled False Positives (top left), which has an opposite polarity to that of the Loss of Trust and Loss of Usability loops just discussed. A false positive, as mentioned in Chapter 1, is a scientific term that means marking something as positive in a test when it should really have been marked as negative. In relation to the spam problem and content filtering, this means filtering non-spam, i.e. wanted, messages. As mentioned
before, false positives directly translate into lost revenue for commercial email and can cause personal problems for personal email.

Additionally, as shown in Chapter 1, false positives are a fundamental problem with applying a spam definition in an ex ante fashion, e.g. content filtering. That is, this loop cannot be erased while content filtering is in place. Doing so would require an ideal definition of spam that we have already shown is unachievable. Therefore, to the extent that false positives are created by content filtering, we expect a decrease in the Expectation Email Delivered variable because non-spam senders are now worried that their email may become false positives. As a result, we expect Non-Spam Sent to decrease, which counteracts our original decrease in the Spam Ratio. Of course, this effect can be mitigated by reducing false positives, or more accurately, the perceived rate of false positives.

The second side effect depicted (bottom) is a loop identified as Arms Race. As content filtering technology has progressed, ways to get around the content filters have progressed in lock-step. There are countless ways to “trick” the filters that are always changing, which is one driver behind the ever-changing nature of spam described in Chapter 1. Some current ways to trick content filters are misspelling words, using random words or phrases, pretending to be random people, and sending email with just an image.

This Arms Race effect is another fundamental problem with content filtering because content filtering provides both incentives to beat the filters and the opportunity to do so. The incentives are obvious; the opportunity exists because one can easily test whether messages are being filtered or not (by sending email to test accounts), and then change the messages in question until they are no longer being filtered.

Moreover, this Arms Race is simply exacerbated by the low costs of sending email mentioned earlier. As more and more spam messages are filtered, at the same time, more and more spam messages are sent. Since the costs are so low, it is really no problem for senders to buy a few more computers and send a few more million messages a day. This is one of the reasons why I believe that spam senders often send what appears to be the same message several times in close time proximity—the messages may actually be slightly different in one way or another (at the very least in time) and the sender hopes that one or more of them will get through the content filtering processes in place.

The third side effect (right) is positive, however. It is labeled Stratification because over time content filtering can have the effect of stratifying spam senders from non-spam senders. That is, the Arms Race is driven by attempts to bypass the solution, i.e. the content filters. These attempts to bypass cause spam senders to change the messages they send. Yet to the extent that false positives are kept to a minimum, we expect only spam senders, i.e. those filtered, to engage in this behavior. Over time, this leads to an increase in the Hardening of Spam Definition variable because the spam senders use tactics that non-spam senders do not. In response, those tactics can become new rules for the content filters, which reinforce their effectiveness.

As you can see, pulling Lever A via content filters creates a myriad of effects on the Spam System. We can and have used the model to trace out these effects. In terms of evaluation, it is unclear whether a given content filter or content filtering in general is good for the spam
problem. To answer that question, one would need to attempt to quantify the effects and compare them. However, I argue that content filtering should be avoided nevertheless.

The reason I take this stance is because other solution categories are more clearly beneficial, which we will see shortly. In particular, other categories do not have the negative side effects of the False Positives and Arms Race loops. These loops, in my opinion, are extremely dangerous and should be avoided if at all possible. Because other solution categories exist that do not have them, such avoidance is possible.

Additionally, as noted in Chapter 3, the content filtering category contains the most widely deployed solutions to date, which have not yet resulted in a useful decrease in the Spam Ratio. Given the rapid changing nature of spam, I infer that the Arms Race loop is dominating the dynamics, negating the positive effects from the other loops. The False Positives loop seemingly has not had that much of an impact to date, but I still believe it is similarly dangerous and could cause significant problems if it were to get out of control. It is particularly dangerous because it is a reinforcing loop as described above.

5.3 Access Filtering

The second category of spam solution types is Access Filtering. As noted in Chapter 3, the major difference between access filtering and content filtering is that access filtering attempts to block alleged spam senders from sending spam in the first place whereas content filtering attempts to block alleged spam after it is sent. That is access filtering literally attempts to block access, and from this perspective, it is easy to see why it pulls Lever B in attempt to directly influence the Access to Methods variable. Again, I will briefly review how each of the particular spam solution activities in this category goes about this task.

- **Best Practices** work to cut off a variety of methods used to send spam, e.g. open relays.
- **Postage Costs** cut off access to low-cost sending for some recipients.
- **Computational Costs** similarly cut off access to low-cost sending for some recipients.
- **Server White Lists** increase access to email servers for certain senders.
- **Server Black Lists** cut off access to email servers for certain senders.
- **The Do-Not-Email-Registry** cuts off access to those emails on the list.
- **Reverse DNS** cuts off access to email servers to those connecting from IP addresses without Reverse DNS in place.

When this category of solutions is operating effectively, Lever B is pulled to achieve an initial decrease in the Access to Methods variable (see Figure 17). As this occurs, Spam Capacity is reduced in turn, which then reduces Spam Sent. And as we have noted several times, as Spam Sent decreases, so does the Spam Ratio in turn.
Therefore, as far as the Base Spam System is concerned, access filtering has a positive effect, namely decreasing the Spam Ratio in the desired direction. However, there are two negative side effects that can result as well and need to be considered (see Figure 18).
These side effects are two loops identified with the same names as was seen in content filtering because the underlying concepts are the same. However, the details are slightly different in this context, and require further explanation. First, the False Positives loop in access filtering stems from declaring non-spam senders as spam senders (as opposed to non-spam messages as spam messages for Content Filtering). This results in increasing the Block Valid Access variable as non-spam senders are blocked from sending.

An increase in the Block Valid Access variable decreases Non-Spam sent as the legitimate senders blocked are unable to send mail. And to the extent that these valid senders make their voices heard, it decreases the ability to do Access Filtering, and completes the loop.

Why this False Positives loop is significantly different from the similar loop in content filtering is that in this context the loop does not reinforce the reduction in Non-Spam Sent directly, but instead balances the application of the method itself. In other words, the loop is not as dangerous because it does not fuel a reinforcing loop where Non-Spam Sent can decrease on its own. Instead, Non-Spam Sent is reduced as an off-shoot of the False Positives loop that reduces the effectiveness of access filtering.

Nevertheless, this False Positives loop is fundamental to discriminatory access filtering in the same manner that the False Positives interaction is fundamental to content filtering. No matter what discriminatory access filtering system one constructs, it will not be perfect, and the False Positives problem will be present in one form or another. This negative side effect can be escaped, however, if the access filtering is not discriminatory. In this case, it is not
really filtering because it applies to all potential senders, i.e. no filtering is taking place. An example would be blocking port 25 (usually SMTP) for all customers. Yet as soon as one does not block for a certain subset of customers, e.g. low-volume senders, the False Positives loop arises.

The second side effect, the loop labeled Arms Race, is another fundamental problem for access filtering, again analogous to the same loop in content filtering. In the case of access filtering, though, the Arms Race creates incentives to increase R&D (as opposed to Spam Sent directly). As Access to Methods decreases, spam senders would of course like to maintain their spam sending rates in order to maintain their Profits. Therefore, they must expend more profits to research new methods in order to bypass those methods that are now blocked by the access filtering.

Again, the Arms Race loop in the access filtering context is less dangerous than that in the content filtering context. This difference stems from the time it takes to respond to the solution. As noted in the last section, content filtering provided both the incentives and opportunity to bypass the solution. The same is true for access filtering. However, in this case the opportunity is mitigated by the significant time R&D takes to develop new methods, while in the content filtering case, bypassing the solution is as easy as sending more spam through existing methods.

The Arms Race loop in access filtering is further mitigated by the use of Profits to access old methods. In other words, access filtering, in the aggregate, decreases Access to Methods by increasing the costs associated with using those methods. In so doing, spam senders have a choice—they can either bear this increased cost or spend that money on R&D to develop other low-cost or similarly low-cost methods. In practice, it is probably a combination. At a low level of increased cost, spam senders might simply pay to keep using the same methods. But as this cost increases, they would invest in new methods as to protect the survival of their business. In any case, I believe the incentives to counteract access filtering are sufficiently in place as to render this Arms Race interaction a fundamental problem.

With regards to evaluation, it should be clear that non-discriminatory access filtering has one positive effect on the spam system. That is, blocking methods for all senders regardless of whether they are sending spam or not causes a reduction in Spam Capacity without a False Positives problem. A variation of this is the Identifiability category discussed in the next section. Yet, for the most part, access filtering is discriminatory and thus suffers from the False Positives loop. And access filtering always suffers from the Arms Race loop.

Nevertheless, these loops are less dangerous than the analogous loops found in content filtering. As such, access filtering should be preferred in general to content filtering. However, I would also argue, like for content filtering, there are still other categories of solutions where limited resources are better spent that do not have significant negative side effects. One such category will be discussed next.
The third category of spam solutions is Identifiability. Both of the spam solutions associated with this category, IP Address Authentication and Content Signing, pull both Levers B and F (see Figure 20). First, by pulling Lever B, they have an intended effect of decreasing the Access to Methods variable, which has the same positive effect on the Spam Ratio variable as outlined in the last section. In particular, identifiability solutions cut off Access to Methods that forge sender information within messages. Recall that they do this by comparing message sender information to that which real domain owners publish in DNS. If the message information matches what is in DNS, then the message passes, and if not, the message may be considered suspect.

Yet Identifiability solutions also pull Lever F, which results in directly increasing Enforcement Capacity. This effect results from the fact that each of these solutions works to help better identify the senders of messages, and hence the category title Identifiability. Such identification is essential to enforcement, both public and private, and thus directly increases Enforcement Capacity. An increase in Enforcement Capacity leads to an increase in Enforcement Success, which in turn increases Perceived Consequences. As we have seen before, an increase in Perceived Consequences leads to a decrease in the Desire to Spam Variable, which decreases Spam Sent in the desired direction.
However, with these positive effects, as we have seen for the other categories, comes at least one negative effect (see Figure 20). In this case, forcing people to use Identifiability solutions has a cost, which is captured in the Cost of Compliance variable in the figure. When you institute any of these solutions, you are essentially raising the cost of sending email for all senders. In the aggregate, we can expect less Non-Spam Sent as a result of an Increase in the Cost of Compliance variable.

Figure 20  Identifiability Side Effects

Yet this negative side effect can be mitigated by working to keep Cost of Compliance minimal. Additionally, it does not create a loop whereby the solution itself is balanced or a negative effect is reinforced. That is, this side effect seems less of a problem than the False Positives and Arms Race side effects created by content filtering and access filtering.

As you can see, those dangerous loops are not present in the Identifiability case. This is because these solutions apply to all email users, as opposed to just some email users determined on a discriminatory basis. That is, as we noted with non-discriminatory access

---

26 For example, solutions could take a phase-in approach where users could gradually opt-in to the solutions over time, minimizing the coordination compliance costs of switching to a given solution at one point in time. Another example is solutions could re-use existing technology already deployed (e.g. DNS), minimizing the training and re-training costs needed for implementation.
filtering, since one is not choosing which senders the solution applies to, there is fundamentally no False Positives problem.

Additionally, Identifiability solutions bypass the other fundamental problem of the Arms Race seen earlier. To see this, consider implementing IP Address Authentication across the board for all email providers. Then there is really no getting around this method for sending email since it is now the new way to send email, i.e. there is no Arms Race. That is, one has changed the basic way to send email to include the solution such that no new methods can be created to avoid it. And since all senders are included, there is no possibility of false positives.

That being said, for a given spam solution, the Cost of Compliance might be so high as to not be worth it. Additionally, adding identity to e-mail may not be desired as well. Nevertheless, of the three solution categories presented thus far, Identifiability solutions seem to deserve significant attention. They warrant such consideration because it is clear from the model that they have positive effects and that their negative effects are limited. Furthermore, these solutions deal with the spam system in a holistic manner by attacking both the legal and technical aspects of the problem simultaneously. In other words, they make technical changes that are useful legally. That is not to say they will completely solve the spam problem alone, but they should be considered as part of general strategies. Such general strategy will be explored in the next chapter.

5.5 Regulation

The effects created by the spam solution category of Regulation are rather simple (see Figure 21). This category pulls Lever B in order to increase the Regulation variable. Solution types in this category include State Spam Protection Laws, Federal Spam Protection Laws, as well as private Acceptable Use Policies (AUPs).
An increase in Regulation leads to additional Enforcement Capacity, i.e. new regulations under which enforcement can usefully be brought to bear. As we saw in the last section, an increase in Enforcement Capacity leads to a decrease in the Spam Ratio via an increase in Perceived Consequences. Mitigating the initial increase in Regulation is the Legal Protections variable.

In Figure 21, there is also a negative effect depicted, namely an increase in the Move to Less Regulated Areas variable. This effect was explored in the last chapter as a balancing effect against an increase in Perceived Consequences via Enforcement Success. In terms of evaluation, because of this negative effect, Regulation is not strictly a positive influence on the spam system. Nevertheless, I believe it is a necessary step to pursue.

Without regulation, the regulative balancing loops in the Base Spam System are ineffective. And, as we have seen in the last section, the Regulative Reaction loop is important in the usefulness of the Identifiability solutions. Without useful regulations present, then the evidence gained from the Identifiability solutions is useless. Additionally, the negative effect created by the International Problem loop can be mitigated through the International Collaboration category presented next. As such, I believe regulation should be pursued to the extent necessary to help other effective solutions. That is, regulation for its own sake is probably unnecessary, but regulation tailored to aid other solutions, e.g. Identifiability solutions, is particularly useful.
International Collaboration as a spam solution category is quite similar to Regulation. In fact, the only difference is that the regulation created is international in nature, as opposed to just applicable to one jurisdiction. As such, the effects produced by International Collaboration are quite similar (see Figure 22).

Figure 22  International Collaboration Effects

Of course, International Collaboration produces regulation, i.e. pulls Lever G to increase the Regulation variable. This increase may not necessarily be binding international law, but could be shared understandings in the form of treaty. For example, a treaty could allow for reciprocal extradition of spam senders. In any case, whatever the form of the increase in Regulation, an increase in Enforcement Capacity results, which leads eventually to a decrease in the Spam Ratio as described in the last and other sections.

The significant difference between the International Collaboration category and the Regulation category is that the negative effect for the Regulation category (an increase in the Move to Less Regulated Area variable) has reversed polarity and been turned into a positive effect for the international collaboration category. That is, international collaboration also pulls Lever E to achieve a direct decrease in the Move to Less Regulated Areas variable. This switch results from the fact that with true international collaboration, there are fewer
places to move to, i.e. less places where those new regulations do not apply. In the case of ideal international collaboration there is almost no where to go to escape the regulation.

What the model does not depict, however, is how hard international collaboration is to achieve. I touched on this earlier in Chapter 3, where I described the slow UN process that is taking place with regards to international standards for information technology related crime. I just note here that achieving true international collaboration is difficult, but some international collaboration is better than none. If a few countries sign a treaty, there are, as noted, then a few less places where violations can easily take place. Therefore, to the extent possible, the model recommends international collaboration.

5.7 Enforcement Capacity

The final and sixth listed spam solution category is enforcement capacity. This category includes spam solution activity that pulls Lever F in attempt to directly increase the Enforcement Capacity variable (see Figure 23). Lawsuits increase enforcement capacity by increasing the enforcement of existing regulations. The proposed bounty system accomplishes this effect by helping prosecutors identify those which they would prosecute if only they could do the identification themselves. And vigilante efforts produce a similar result.
As we have seen in the past two sections, an increase in Enforcement Capacity can lead to a decrease in the Spam Ratio as desired. Certainly directly increasing Enforcement Capacity by pulling Lever F also leads to this effect. Also, an increase in Enforcement Capacity can fuel further increases in Enforcement Capacity via the Precedent loop explored in the last chapter.

For example, bringing lawsuits successfully can make it easier to bring more successful lawsuits. There are at least three reasons behind the Precedent effect in this case: (1) often the same legal teams bring suit, and thus their learned processes can be reused over time; (2) similarly, the legal record of suits can be used by other legal teams; and (3) suits result in court precedents that are reused by judges over time.

The negative interaction associated with increasing Enforcement Capacity is the same as that associated with Regulation, namely an increase in the Move to Regulated Areas variable via the International Problem loop. The reasons are also the same, but perhaps more clear in this case. For example, if suits are being actively brought against alleged spammers, and there are bounties out for them, it is not too hard to see why they might move to another country where those things are less present.

Therefore, in terms of evaluation, enforcement capacity solutions are neither strictly bad nor good, and depend on the tradeoff between the increase in Perceived Consequences from Enforcement Success, and the decrease in Enforcement Capacity due to the International Problem. As such, I believe the model derives the same advice as it did for the Regulation category, namely Enforcement Capacity should be pursued in a targeted manner where necessary to aid other effective solutions, e.g. International Collaboration or Identifiability.

5.8 Spam Solution Interactions Summarized

In this chapter, we have systematically used the model of the Base Spam System to qualitatively describe the effects of spam solutions to date. Each category represents a set of different levers pulled by spam solutions. We have seen that most categories produce negative effects in addition to the positive effects sought. Some of the effects have been noted as stronger as others. For instance, the False Positives and Arms Race loops created by Content Filtering are particularly dangerous, whereas the Cost of Compliance effect created by Identifiability solutions can be mitigated through careful planning.

Throughout the chapter, I have noted preferences for certain categories as a result, namely Identifiability over Access Filtering over Content Filtering. I have also noted that Regulation and Enforcement Capacity can be used to help other solutions, e.g. Identifiability, in addition to being pursued alone. In general, solutions are not mutually exclusive, and you can see from Chapter 3, most are going on presently in a simultaneous manner. In the next chapter, I will attempt to synthesize the discussion in this and previous chapters in an attempt to give advice to those trying to form effective spam solution strategies.
6 Forming Effective Spam Solution Strategies

In the previous chapter, the spam solution activity presented in chapter 3 was evaluated using the model of the Base Spam System outlined in chapter 4. In so doing, we revealed how the six categories of spam solution activity identified interact with the spam problem. Negative side-effects in addition to the desired positive effects sought were identified.

In this chapter, we synthesize the previous discussion and explore the topic of forming effective spam solution strategies. By pluralizing strategy I mean to emphasize that there may be more than one effective spam solution strategy. As such, I hope to offer a general way of thinking about spam solution strategy that can be used to advocate for more specific strategies.

6.1 Key Spam Solution Issues

First, I would like to highlight the following issues associated with the spam problem that I believe to be key issues when thinking about forming effective spam solution strategies. These issues have all been mentioned in previous chapters.

- **Precise Spam Definition**: one that would declare sharp rules that could be used on any email message to tell if it were spam or not. That is, \textit{ex ante} application of the definition would not result in a category of messages that require further investigation.
- **Accurate Spam Definition**: one that if used and if it declared a message as spam, it would indeed be spam.
- **Response Subset**: the existence of some real subset of the general population that buys products and services advertised from spam.
- **Low Monetary Cost**: the ability to send millions of emails a day at relatively low monetary cost.
- **Low Computational Cost**: the ability to send millions of emails a day at relatively low computational cost, e.g. it does not take much time or resources to send an email such that one can use relatively cheap hardware to do so.
- **False Positives**: declaring messages are spam when they are not spam, or declaring senders as sending spam when they are not sending spam.
- **False Negatives**: declaring messages as not spam when they are spam, or declaring senders as not sending spam when they are sending spam.
- **Arms Race**: engaging in a back and forth escalating process with those who intentionally send spam because one provided them the incentives and opportunity to do so.
- **Identification**: figuring out who exactly sent an email. This process is difficult now because of the technical realities of the SMTP protocol.
- **Legal Protection**: tailoring regulation to not infringe on legal rights so much so to make them unenforceable.
- **International Problem**: controlling those who intentionally send spam messages across jurisdictions.
- **Stratification**: spam senders self selecting themselves through behavior.
First, recall the set of levers within the Base Spam System model as presented in the last chapter (see Figure 24). These levers are also listed in the following table (Table 6) along with the spam solution categories that attempt to pull them appropriately.

**Table 6**  
**Base Spam System Levers and Spam Solution Categories**

<table>
<thead>
<tr>
<th>Lever</th>
<th>Spam Solution Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lever A: Spam Ratio</td>
<td>Content Filtering</td>
</tr>
<tr>
<td>Lever B: Access to Methods</td>
<td>Access Filtering, Identifiability</td>
</tr>
<tr>
<td>Lever C: Response Rate</td>
<td></td>
</tr>
<tr>
<td>Lever D: Perceived Consequences</td>
<td></td>
</tr>
<tr>
<td>Lever E: Move to Less Regulated Areas</td>
<td>International Collaboration</td>
</tr>
<tr>
<td>Lever F: Enforcement Capacity</td>
<td>Enforcement Capacity, Identifiability</td>
</tr>
<tr>
<td>Lever G: Regulation</td>
<td>Regulation</td>
</tr>
<tr>
<td>Lever H: Recipient Concern</td>
<td></td>
</tr>
</tbody>
</table>

From this table, you can see that most of the levers identified are pulled by at least one spam solution category already discussed. The levers not targeted by at least one spam solution category already discussed are Lever C (Response Rate), Lever D (Perceived Consequences), and Lever H (Recipient Concern). Note, however, as was explored in the last chapter, the
variables associated with these levers are indirectly influenced by some of the other spam solution categories. They are just not directly targeted by those spam solution categories.

Nevertheless, I believe two new categories can be created to target these levers directly: education and publicity. First, consider the new category of education. This category pulls both levers C and H in the desired manner (see Figure 25).

**Figure 25** **Education Effects**

Education in this context would involve efforts to educate the public at large about the spam problem, including what they could do to help solve it. Such efforts could target Recipient Concern, Response Rate, or both, and by targeting one, the other is likely to feel at least some ancillary effects. For example, one could try to educate email consumers that buying products and services from spam hurts all email users, analogous to anti-drug advertisements linking drug money to terrorism. Or one could simply raise awareness of the spam problem by highlighting through advertising that individual email consumers can help the spam problem by calling their congressmen and stating their interest.

If, in the aggregate, Recipient Concern is increased, then the model predicts Spammer Alienation to increase from what it would otherwise be and Regulation to eventually increase from what it would otherwise be as well. Similarly, if, in the aggregate, Response Rate were to decrease as desired, we can expect Responses to decrease from what they would otherwise be, and in turn Profits, which as we know plays a central role in the Ability to Spam, Motivation, Innovation, and Initiative loops.
To my knowledge, specific large-scale efforts that would fall into this education category do not exist. There are smaller efforts in the form of Web sites that help to educate the public, however. The main problem with this spam solution category in practice is that it is hard to measure the effects of particular campaigns, and thus it is hard to predict a return on investment to compare with other possible efforts. A second related problem is that the target of such education efforts would have to be all email users, and usefully reaching that large group is difficult and expensive. One way of course would be to use spam to educate about spam, but that process seems like it would be setting the wrong example coming from those who prosecute spam senders, i.e. the government.

The second new spam solution category is entitled publicity and pulls lever D in an attempt to directly influence the Perceived Consequences variable (see Figure 26). If an increase in that variable is achieved, we expect a decrease in the Desire to Spam variable. Such a decrease, as we saw in the last chapter, should influence Spam Sent directly in the desired direction as well as indirectly in the same direction via the Initiative loop. At the same time, however, we expect an increase in Perceived Consequences to lead to a reactive increase in the Move to Less Regulated Areas variable.

Figure 26 Publicity Effects
Solutions in this publicity category would attempt to publicize successful enforcement of spam senders. Presently, we see this in an ad hoc fashion via media outlets picking up stories of spam lawsuits. However, there does not seem to be a widespread targeted campaign to catalogue enforcement success and publicize it appropriately. Doing so seems less daunting than efforts in the previous education category because the group one is trying to reach (spam senders) is orders of magnitude fewer people. Additionally, this group is presumed to be a relatively tight knit community, and so reaching them via spam sending forums, computer magazines, common Web sites, etc. should not be as difficult (or costly). In the next section, we consider these two new categories with the six other categories already explored and make some initial strategy observations.

6.3 Initial Strategy Observations

Now that we have a more complete picture of how one could engage the possible levers, let us turn to the negative side effects created by such engagement. The following table (Table 7) lists the now eight spam solution categories and identifies the levers targeted as well as the negative side effects produced.

<table>
<thead>
<tr>
<th>Category</th>
<th>Lever(s)</th>
<th>Negative Side Effect(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content Filtering</td>
<td>Lever A: Spam Ratio</td>
<td>Arms Race False Positives</td>
</tr>
<tr>
<td>Access Filtering</td>
<td>Lever B: Access to Methods</td>
<td>Arms Race False Positives</td>
</tr>
<tr>
<td>Identifiability</td>
<td>Lever B: Access to Methods</td>
<td>Cost of Compliance</td>
</tr>
<tr>
<td></td>
<td>Lever F: Enforcement Capacity</td>
<td></td>
</tr>
<tr>
<td>Regulation</td>
<td>Lever G: Regulation</td>
<td>International Problem</td>
</tr>
<tr>
<td>International Collaboration</td>
<td>Lever E: Move to Less Regulated Areas</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lever G: Regulation</td>
<td></td>
</tr>
<tr>
<td>Enforcement Capacity</td>
<td>Lever F: Enforcement Capacity</td>
<td>International Problem</td>
</tr>
<tr>
<td>Education</td>
<td>Lever C: Response Rate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lever H: Recipient Concern</td>
<td></td>
</tr>
<tr>
<td>Publicity</td>
<td>Lever D: Perceived Consequences</td>
<td>International Problem</td>
</tr>
</tbody>
</table>

It is tempting to simply discard those categories with significant negative side effects in favor of those without significant negative side effects. However, doing so would be premature for two reasons:

1. Just because negative side effects exist does not mean that the effort as a whole is not net beneficial for the spam problem. For instance, recall that content filtering can have a positive stratification effect.
2. As noted in the last chapter, the categories are not mutually exclusive. We must recognize that the spam problem is really a complex system, namely the Base Spam...
System presented in Chapter 4. As such, just concentrating on one category may have the desired effect on the levers pulled in that category, but may leave open or create ancillary effects on the other variables.

For example, one might declare after looking at this table that education and international collaboration should be only pursued. Even if those efforts are successful somewhat, Access to Methods might still be so high to keep the Spam Ratio significant. In other words, the first initial observation we can make about forming effective spam solution strategies is that a holistic approach is necessary. When forming strategies, we need to keep in mind the entire Base Spam System and make sure we are plugging all the holes, so to speak.

That being said, where the same general effect can be achieved through multiple categories, one should avoid the categories with negative side effects to the extent possible. That is, one should avoid unnecessary negative side effects, which is the second initial observation we can make about forming effective spam solution strategies.

Third, in the broadest sense, we want to avoid activity that has a net neutral or net negative effect on the spam problem. That is, on net, resources spent towards that activity are wasted. The reason why net neutral would still be bad is because we are assuming those resources could have been spent to attain net positive results, and thus, if resources are spent towards a net neutral activity, we have lost the opportunity cost from those expenditures.

More specifically, net neutral and net negative effects result when the desired net positive effect is not achieved because either the particular solution in question is implemented poorly or because the system reacts to counteract the desired effect. We shall ignore the former in hope that particular solutions are well-implemented. The latter happening occurs when creating the solution also provides the incentives and opportunity to counteract it to the degree that counteracting efforts dominate the dynamics. In the next section, we consider these initial observations in the formulation of a general spam solution strategy.

6.4 General Strategy

The previous three sections have synthesized much discussion about the spam problem and the results obtained from the model of the Base Spam System. We have seen the key spam problem issues and the key insights from the model. However, what has not been done is a quantitative examination of particular solutions or categories, which is beyond the scope of this thesis.

As such, the model does not necessarily recommend one particular strategy over another at this time. Nevertheless, I believe the insights gained present a persuasive argument to move in a general direction with regards to forming an effective spam solution strategy. I will now present this argument and then an example strategy in the next section, followed by a stakeholder analysis of that example strategy and an overview of its difficulties.

I have already argued that the negative side effects labeled Arms Race and False Positives are particularly dangerous. I have also shown that these effects are fundamental problems.
associated with both content filtering and access filtering. As we have seen, however, both problems are less concerning with regards to access filtering. Additionally, the False Positives problem has been seemingly mitigated to date.

On the other hand, given that the spam problem has not subsided in the face of significant content filtering and access filtering, I infer that the Arms Race problem is dominating the dynamics of those solution categories. Further indication of such domination is the rapidly changing nature of spam and the rapid introduction of new spam methods, both of which are evidence of spam senders bypassing solutions in an Arms Race fashion.

If there were no other way to control Access to Methods than content filtering and access filtering, then I believe they would need to be a part of an effective strategy. However, there is another solution category that potentially has similar effects but does not have these intrinsic side effects, namely identifiability. These solutions fundamentally change email so that Access to Methods used to hide and steal identity is reduced for all email users. As we have seen, because the email protocols themselves are changed, there is no Arms Race or False Positives problem. The drawback of these methods is the Cost of Compliance. However, the current leading proposals I believe do a good job keeping this cost under control by phasing in the changes over time.

Additionally, I would advocate non-discriminatory access filtering, which as mentioned previously, is a special case of access filtering that is not usually done. In non-discriminatory access filtering, there is no filtering per se, but instead simply a ceasing to allow use of particular methods, e.g. open proxies and open relays. As we have noted, access filtering in general has the negative side effects mentioned, but if it is non-discriminatory, i.e. applies to all email users like identifiability solutions, it escapes the False Positives problem. And coupled with identifiability solutions that hinder identity masking, the Arms Race problem would be mitigated because further identity masking methods would be harder to come by.

As discussed in the last chapter, the identifiability solutions attack the problem holistically by producing technical change at the same time as increasing Enforcement Capacity. To aid in that increase in Enforcement Capacity, the categories of regulation and enforcement capacity should also be sought to the extent necessary to aid the identifiability solutions. How this might be done is explored in the example strategy in the next section. Basically, though, it involves creating regulation and providing enforcement resources necessary to get significant enforcement success.

With identifiability coupled with regulation and enforcement capacity, one has effectively made spam senders identify themselves and face enforcement or move to less regulated areas. To the extent that the latter occurs, the international collaboration category should be pursued to mitigate this negative side effect.

What we have not talked about yet in this general strategy discussion is the new categories of education and publicity. I do not think education is effort well spent because I believe that most email users are aware of the spam problem. However, one could test whether such effort is useful through surveys. Similarly, it is unclear whether publicity is useful because one might assume spam senders monitor enforcement success closely. Nevertheless, I think it is a useful addition to general strategy because it is relatively inexpensive.
To summarize this section on general strategy, I have argued that the current major expenditures used for content filtering and access filtering would be better re-appropriated towards identifiability and enforcement capacity. Doing so avoids significant negative side effects that seem to be presently dominating the spam system dynamics and preventing the alleviation of the spam problem. However, as noted in a subsequent section, this may not be likely, though identifiability can be sought independently.

Additionally, this general strategy takes a holistic view of the complex system that is the spam problem by jointly tackling the technical and legal issues. To the extent that this strategy is undermined by the Move to Less Regulated Areas and Perceived Consequences variables, international collaboration and publicity should be sought to counterbalance. Similarly, significant regulation needs to be in place to enable successful enforcement. In the next section, I will present an example strategy using this logic.

6.5 Example Strategy

This section enumerates an example strategy using the general strategy discussion from the previous section as a basis. I would also like to reiterate that there may be several effective spam solution strategies. First, consider the following facts discussed previously that help us funnel the general strategy into the example strategy.

- There are only a few hundred major spam operations.
- A handful of email service providers, e.g. Hotmail (Microsoft), Yahoo!, AOL, etc., provide a large share of email accounts throughout the Internet.
- Most people who use the Internet use email, which includes most of the adult population of the US.

Given the CAN-SPAM Act is in effect, and, as mentioned, is the result of increasing Recipient Concern over several years that also resulted in many US state laws, we can conclude two things: (1) Recipient Concern is significant enough to have already resulted in significant Regulation; and (2) given this result, Recipient Concern must be relatively high. Additionally, although many press reports have painted a negative picture surrounding this law because of no traceable net-positive result to date, the law has not really been tested in court in any significant way. That is, there has not yet been enough significant Enforcement Success. Until such Enforcement Success occurs, the model predicts that we should not expect a significant increase in Perceived Consequences and, in turn, a decrease in Desire to Spam.

With these points in mind, consider the following example strategy.

1. Agree on a spam definition.
2. Enact regulations prohibiting spam according to the agreed definition.
3. Enact technological change to only the extent necessary to enable the introduced legislation realistically enforceable.
4. Identify major spam senders that can be prosecuted under the law.
5. Promptly and strictly enforce the introduced legislation, and in so doing, permanently shut down operations sued.
6. Publicize successful enforcement.
7. Engage in international treaty to mitigate moves to less regulated areas.

The interaction of these steps with the Base Spam System is highlighted in Figure 27.

Figure 27  Example Strategy

The goal of this strategy is to make the Perceived Consequences variable high enough that spam senders believe that they can expect to be promptly shut down if they engage in spam sending. Additionally, the strategy simultaneously reduces Access to Methods to the extent necessary to enable enforcement and keep the Perceived Consequences variable high. Note that these steps would be conducted in an iterative process. That is, if on the way to step 5, it is discovered there is problem with a previous step that makes it hard to enforce the regulation from step 2, we need to back up and either do more of steps 2 or 3 to enable both steps 4 and 5. Now I will briefly consider each step in a little more detail.

In terms of the spam definition (step 1), it is important to agree on a definition because if one dominant definition of spam could emerge, activity currently spent arguing about what is spam and what is not spam could be refocused on more proactively solving the spam problem. Additionally, the choice of the definition of spam can shape the efficacy of spam solution activity, and thus have a pronounced impact on the timeliness of solving the spam problem. As such, I believe the definition that should be chosen is that adopted by this thesis and the CAN-SPAM Act of 2003, namely *email that uses fraudulent tactics or does not contain a valid opt-out mechanism.* In Chapter 1, Section 9 (page 16), I enumerated several reasons why
this choice is particularly useful. One major reason is that it stratifies spam senders from non-spam senders as mentioned.

In terms of regulation (step 2), the CAN-SPAM Act of 2003 may be enough public legislation because (a) this Act enables the government to prosecute fraudulent spam senders; and (b) creates significant penalties for violation. However, it may not be enough if it proves too difficult to enforce, which only time will tell. Additionally, it may not be enough if after step 7 there is still enough non-fraudulent spam to keep the Spam Ratio high. If that is the case, then further regulation would need to be enacted to target a significant portion of the remaining group of spammers, i.e. the definition of spam would have to be expanded.

If we are to go with the CAN-SPAM Act as our primary regulation for implementing this strategy, then individual states should concentrate on helping the Federal government enforce that law as opposed to any state laws that may exist. The CAN-SPAM Act is much more equipped to handle jurisdictional issues within the US, and is likely to be constitutionally valid. Additionally, the Act enables a central enforcement agency, namely the FTC, to act as a central repository for information necessary to prosecute targeted operations. Instead of 50 states acting independently, all can now pool their resources by sending useful information to this agency. Finally, the FTC with the backing of the Federal government is in a better position to deal with the international nature of the problem.

In terms of technical change (step 3), the example strategy proposes that we should concentrate, technically, on the issues that make step 4 (identification) and step 5 (enforcement) difficult. In terms of the spam solution categories, and as we saw in the last section, this goal aligns perfectly with the identifiability category. In other words, what we need is a way to show that certain messages were sent by certain operations to or from certain places.

There are potentially many technical ways to do this, including the IP Address Authentication and Content Signing proposals already mentioned. Since at least IP Address Authentication seems to be happening, we can wait to see if that activity will have the desired effect. The desired effect is of course prohibiting Access to Methods that allow the masking of identity such that spam senders are forced to provide some traceable information. This information can be used to do identification (step 4) and then enforcement (step 5). In addition to identifiability, non-discriminatory access filtering may aid these goals, e.g. prohibiting open proxies and open relays as mentioned in the last section.

In terms of identification (step 4), vigilante groups such as the Spamhaus Project already do a good job of identifying spam operations. Since there are so few major operations, after technical change produces useful traceable information, it should be additionally relatively cheap to physically find these people. The FTC spam group should be tasked with this job since they are also doing the enforcement. The main problem with this step, then, is tying particular messages to particular people.

Again, with the technological changes enacted in step 3, this process should become much easier. Vigilante groups are in a position, already maintaining data on spam operations, to
simply add more data to their databases. Therefore, channels between these groups and the big email providers should be instituted and/or become more open, with appropriate data flowing between the providers and the groups. Similar channels, to the extent they do not already exist, can then be established between the vigilante groups and the FTC or other enforcement agencies.

Once found and identified definitively as spam senders, the actual people behind these operations should be punished enough to deter other people from taking their place (step 5). I imagine this means that these people would have to face some sort of jail time via criminal statutes because civil fines will probably not be enough given widespread bankruptcy laws and other legal avoidance mechanisms.

The first major hurdle to enforcement is testing the existing law in court. It might have been drawn to narrowly to enforce usefully, or too broad to pass constitutional muster, though I personally doubt both. If it proves to be unenforceable, it needs to be changed by Congress, so that useful enforcement can proceed as quickly as possible.

After success, such success should be publicized as much as possible (step 6). Pursuing this step involves the publicity category of solutions mentioned earlier in this chapter. In other words, successful enforcement needs to be made known to all spam senders such that the Perceived Consequences variable increases significantly. If for some reason isolated spam senders are unaware of successful enforcement, the strategy is undermined because no voluntary ceasing of spam sending would take place.

Finally, we expect from the model that if the Perceived Consequences variable is raised significantly, some spam senders will move to less regulated areas. Hence the last step (7) involves balancing this International Problem by targeting it directly via the international collaboration solution category.

### 6.6 Example Strategy Stakeholders

To further consider the example strategy outlined, this section identifies how major stakeholders might interact with it. The next section identifies and discusses some key challenges in its implementation.

**ISPs and ESPs**

Perhaps the most critical stakeholder group in terms of implementing the example strategy is comprised of Internet Service Providers (ISPs) and Email Service Providers (ESPs). The reason why this group is so critical is that they are necessary to achieve a critical mass for the identifiability solutions (step 3). These solutions, being technically on top of email, have similar network efforts to email. That is, the more email accounts are using them, the more useful they are. A quick way to achieve critical mass would be for the large ISPs and ESPs to implement compatible solutions. I do not expect a large barrier for such implementation because it is in the interest of ISPs and ESPs to implement identifiability solutions. They are quite cheap to implement and can have a promise to reduce the spam problem for the email accounts they operate, which is obviously good for business. We have already seen such
activity taking place through AOL, Hotmail and Gmail’s implementation of SPF as discussed and through the Anti-Spam Technology Alliance where the major ISPs and ESPs are exploring such solutions.

A potentially bigger challenge with this stakeholder group is their assistance in identifying spam senders (step 4). Controlling most email accounts means that they also are bombarded with a significant amount of spam. Appropriately logging and distilling this information after identifiability solutions have been implemented will be critical to enforcement success (step 5). The challenge is that creating and maintaining the systems to do so may not be seen as the highest priority for this stakeholder group. However, this challenge is mitigated by the private cause of action for ISPs and ESPs available to them under the CAN-SPAM Act. As discussed, major firms have already filed suits under this provision, and as such have identified that collection of evidence is important to them. Therefore, I believe ISPs and ESPs do not present substantial barriers in terms of the example strategy.

State Governments
State governments can similarly help with steps 4 and 5 in the identification and prosecution of spam senders. As mentioned, several states, including New York, Texas, and Washington State have already pursued and prosecuted spam senders successfully. In and of themselves, state governments do not present a significant barrier to implementation because their assistance is not strictly needed. However, they are in a position to help speed up the potential positive effects of the strategy.

State governments via their Attorney Generals’ offices have a unique ear to recipient complaints about spam operations within their states. It is through these offices that the states mentioned have initiated their law suits to date. It has been in their interest to do so because of high recipient concern in their jurisdictions. Nevertheless, as discussed, the CAN-SPAM Act preempted much of their state spam laws. As a result, states may now be less likely to bring similar suits. This would be unfortunate because the resources currently spent on combating the spam problem by states could be refocused on helping the FTC collect evidence as needed and as directed by that agency.

It is understandable though, given federal preemption and FTC oversight that states may cease significant efforts to combat the spam problem. One possibility to help realign incentives would be to provide some funding via the FTC to help state-focused investigations. After all, investigations within a given state would seemingly be better operated by that state given their obvious knowledge advantage about the state environment.

FTC
In the example strategy, the FTC is tasked with orchestrating three critical steps: identification, enforcement, and publicity (4, 5, and 6). This is a logical conclusion given, as discussed previously, the FTC’s forefront role in regulating the Internet to date. However, to be successful in this major role, the FTC will have to devote significant technical and legal resources to these tasks. Doing so could be difficult given that they have other priorities, and must constantly defend their budgetary choices to Congress.

A potential barrier to implementation is that if the strategy starts to work, as predicted by the model, less Recipient Concern could result. Such a decrease in concern could reduce the
political demand for solving this issue, within the FTC and Congress. As a result, the solution strategy could stagnate or reverse if resources are then diverted away from this issue to other issues. One way to mitigate this risk is to, as just discussed in the ISP and ESP subsection, encourage those firms to privately act against spam senders. Then even if FTC incentives for enforcement are reduced, the ISPs and ESPs can maintain pressure since they more immediately feel the effects of the spam problem.

**Congress**

Congress is of course essential in enacting appropriate legislation (step 2). As discussed, their initial job may be complete if the CAN-SPAM Act turns out to be useful enough to carry out the example strategy. If not, this law will need to be revised to enable the successful completion of the strategy. However, revising laws can be a very slow process. Even more so with Congress than with the FTC, if the given issue is not currently politically a major issue, then it might not receive Congress’ attention.

Similarly, Congress is in control of the budget of the FTC. If the spam problem issue wanes with the public, the FTC’s budget allocated for solving the spam problem could be under pressure. Such pressure could be further exacerbated if Congress takes a budget cutting approach in reaction to recent deficits. Again, the private cause of action in the CAN-SPAM Act could mitigate this risk, assuming the CAN-SPAM Act enables successful enforcement.

**Domain Owners**

Like major ISPs and ESPs, domain owners are essential to implementing identifiability solutions to the extent that those solutions center on domains, which they currently do (step 3). Unlike ISPs and ESPs, however, there are not a few domain owners that can achieve critical mass. Instead, most domain owners need to take an affirmative action to make these solutions work effectively, and that could mean actions by hundreds of thousands of organizations. Such a coordinated effort could be difficult, and as discussed, this is the cost of compliance negative side effect of these solutions.

Yet, this coordination problem should not significantly undermine the example strategy because identifiability solutions create incentives for the domain owners to take the action needed. In particular, as long as they perceive that taking such action will increase the probability that their email will be perceived correctly, i.e. not as spam, then it is in their interest to do so. Generally, this is the point of such action, so by default, identifiability solutions provide such incentives. Furthermore, if major ISPs and ESPs were to implement these solutions beforehand, which they are doing, then these incentives are greatly increased. This increase is because those ISPs and ESPs control many accounts and so domain owners can already increase the probability their email is perceived correctly on a high number of email accounts.

**Foreign Countries**

Foreign countries present a unique problem to the implementation of the example strategy. In particular, they are involved in the final step (7) where we try to control spam senders in jurisdictions beyond the scope of US law. This challenge is discussed in the next section, so I will only briefly discuss it here. From a stakeholder perspective, there may be incentives to cooperate or not cooperate depending on the country and the way its citizens use email.
If it is a country is like the US where email is an essential part of commerce and email recipients are heavily targeted by spam senders, we can expect their government to feel similar pressure to the US government. In that case, we can expect their cooperation and probably their action to undertake a similar strategy. If, however, their country does not use email extensively, their citizens are not spam targets, or their citizens derive great benefit from spam sending, they might be less cooperative.

**Recipients**

Recipients, as discussed, drive this problem by being targets to spam, responding to spam, as well as raising concern to solving the spam problem. However, the example strategy does not really involve them in any direct fashion. This is by design since this group is made up of most people and thus has widely divergent characteristics and any solution involving all recipients would face significant coordination problems.

The example strategy would essentially reduce the Spam Ratio for recipients over time, which would presumably be welcomed. One issue already highlighted is that by doing so, Recipient Concern might be significantly reduced, which could result in lower priorities for ISPs, ESPs, Congress and the FTC in dealing with the spam problem. That is not necessarily bad if the problem is solved though. That is, if Recipient Concern is low enough because recipients are not wasting much resources dealing with a small Spam Ratio, then perhaps resources that were spent lowering the Spam Ratio could now be reallocated to other problems.

**Spam Senders**

Spam senders are of course the target of the example strategy. The key desired effect with this important stakeholder group is that an Arms Race feedback loop does not dominate the dynamics of the strategy. As previously discussed, this positive feedback loop can be inferred to be dominating presently. To ensure that it does not dominate the example strategy, we must not provide the capacity and opportunity to easily bypass solutions. The incentives to do will be there because spam senders presumably do not want to be deprived of their current profits.

With identifiability solutions, there is certainly less opportunity and capacity to engage in Arms Race behavior because these solutions apply to all email. Recall that for content and access filtering, the Arms Race was easily achieved because systems could be employed to bypass the solutions by attempting to produce false negatives and false positives. With identifiability solutions, there are no false positives and false negatives. There can still be, however, attempts to bypass the solution. These attempts must be controlled so as not to trigger the Arms Race dynamic, or at least not in the same speed as is currently underway.

Attempts to bypass the solution will arise over time and are difficult to predict for obvious reasons. However, two major possible attempts can be identified. First, to the extent identifiability solutions rely on domains, spam senders can acquire many domains and switch between them as needed. As discussed in the next section, this should create more evidence that can be used for prosecution though, and so may not be desired by spam senders. Second, as discussed previously and also in the next section, spam senders can move overseas to escape US jurisdiction. This particular problem is addressed by step 7 of the strategy via international collaboration.
This stakeholder analysis has shown that the example strategy provides incentives to critical stakeholder groups to carry out the strategy as designed. The next section will discuss some key challenges to doing so.

6.7 Example Strategy Challenges

If the example strategy is net positive, which I believe it would be, then it arguably should be pursued over current strategy that is arguably not net positive given the steadily increasing Spam Ratio over the last decade. However, this example strategy has a couple of key challenges.

First, the solution categories of access filtering and content filtering are likely to continue regardless of any new strategy direction. The reason is that when these categories are isolated from the rest of the spam system they make sense in the marketplace. When people and firms buy filters and firewalls, they are not paying for the negative side effects they create. Instead, they pay for the isolated direct positive effects they create. From an economics standpoint, we can say that both content filtering and access filtering create negative externalities. What further exacerbates this problem is that content filtering and access filtering solutions are the easiest to pursue, and can be made by individual firms and people.

The presence of content filtering and access filtering does not necessarily undermine the example strategy, however. At the very least it means that resources spent towards those spam solutions could be spent more productively to end the spam problem faster. Yet they could undermine the example strategy if they interact negatively with identifiability solutions in the technical change step. For example, we have noted that such identifiability solutions would be best done through a relatively slow phase-in to reduce cost of compliance. If content filters were to act strictly against those not phased-in, many false positives and false negatives may be created in the process, hurting the overall spam problem via an increase in the Spam Ratio. Such an increase could be wrongly attributed to the identifiability solutions and undermine their acceptance.

Second, if there is successful enforcement and people really do go promptly to jail, i.e. enforcement is believable, then it is expected spam sending will largely cease within the US. However, as also mentioned, this would not stop non-US operations or those US operations that feel they are making enough profits to take their businesses over seas. Thus, as noted, it is probably necessary to make some agreements with other countries to prosecute non-US spam senders. That is, engage in some international collaboration.

Note that new binding international laws are probably unnecessary, however. This is because, while spam operations can exist outside the US, we are concerned with those that target the US. Since our laws probably have enough jurisdictional authority to prosecute successfully, the real issue involves extradition—getting the particular people in question back to the US so that they can be brought to justice. And for other countries similarly
interested in solving the spam problem for their citizens, they would simply need similar national laws and push for similar extradition.

The problem with getting extradition and international collaboration in general, however, is that many countries probably do not care about the spam problem and is not a politically important enough issue to the US to press their involvement. That being said, the UN process already mentioned is evidence that there is some concern and political will to act. Given that we might only need extradition agreements, such minor collaboration might be achievable if narrowly constructed.

Another more adversarial way around such collaboration, however, would be to block Internet traffic to and from those countries that do not concede. This is possible to do because it is possible to block traffic based on IP addresses, which can largely be traced to countries, as mentioned previously. Countries certainly do not want their US Internet blocked, and thus most would probably concede under such circumstances. There are of course many issues with this policy, which are beyond the scope of this thesis. I mention it only as a possibility worth consideration elsewhere. (Note that the federal government would not necessarily have to be behind the blocking, but a consortium of large ISPs could presumably undertake it alone.)

Third, even after identifiability solutions are implemented, it might still be prohibitively expensive to physically locate spam senders. What the current identifiability solution proposals do is tie messages to particular domains, and thus spam senders to particular domains. This pushes the identification of spam sender problem to the identification of who operates a given domain. This problem is easier to solve than figuring out who sent a given forged email message, but it still can be difficult.

To procure a domain, payment information needs to be exchanged. Such information is potentially much more useful in terms of traceability than forged email headers, but payment information can of course also be fraudulent in many ways. In any case, it is an empirical question whether one can track down domain owners for a cost low enough to be useful in the enforcement of spam regulations. If it is not possible, there are still options, however. One option is to make domain procurement stronger. Another is that content filtering can be potentially brought back and made more useful through reputation systems based on identifiability solutions. That is, if certain domains are truly identified as spam senders, those domains can be blocked as we are now confident they are not forged. However, doing so brings back the dreaded false positive and arms race problems.

6.8 Summary of Conclusions

In this chapter, I have attempted to synthesize the discussion from the previous chapters in order to inform the making of effective spam solution strategies. Now I would like to conclude this thesis by summarizing that synthesis.
First, the spam problem is really a complex system, and should be dealt with as such. This complex system is made up of variables that are various causes and effects that are connected in ways that create both reinforcing and balancing feedback loops.

Second, one can interact with this complex system by targeting certain variables with spam solution activity. In so doing, one can expect the system to react based on the variable targeted. Such reaction will result in other variables being influenced, which can lead to a net negative or net neutral effect on the system. Solutions that produce such effects should be avoided.

Third, when constructing effective spam solution strategy, a holistic approach needs to be taken that considers the entire system. One example strategy that appears to do this has been outlined, which concentrates on increasing the identifiability of spam messages so that spam senders can be more easily prosecuted under existing law. The holistic approach also necessitates examining the International Problem “hole,” which would need to be plugged in some fashion to be ultimately successful.

Fourth, the goal of any effective spam solution strategy is to reduce the Spam Ratio to a reasonable amount. At that point, the ratio of spam email to all email for most recipients should exist at a reasonable level, and wanted commercial email should be delivered appropriately.

Finally, I believe the reason the suggested example strategy is being pursued less than would otherwise be expected is the very reason that it is effective—it is holistic. It requires significant coordination and alignment between various stakeholder groups (including the Federal government) that has simply not yet taken place. In other words, while it may seem simple because it is easily explained, such a holistic approach involving significant coordination and allocation of government funds is never simple in practice.

This also explains why content filtering and access filtering are so prominent—because they are so easy to pursue. Individuals can produce and engage such solutions in isolation via inexpensive software. With the advent of the CAN-SPAM Act, however, I hope that the necessary coordination to pursue more holistic strategy can begin to form. Even if you do not find my argument for such coordinated strategy particularly persuasive, I hope that this discussion has presented the spam problem in such a way as to enable you to think about the problem more effectively.

America Online, Inc. AOL Junk E-mail Archive: Decisions & Litigation. 2003. 5 July 2004

America Online, Inc. Sender Policy Framework - Wizard. 9 June 2004. 29 June 2004

America Online, Inc. Technical Standards for E-mail Delivery. 2 June 2004. 20 July 2004
<http://postmaster.aol.com/guidelines/standards.html>.

America Online, Inc. Whitelist Information. 2002. 1 July 2004

Associated Press. Gates: Buy stamps to send e-mail. 5 Mar. 2004. 9 July 2004


Brightmail, Inc. Brightmail Anti-Spam 5.5 Data Sheet. 2003. 11 Aug. 2004
<http://www.brightmail.com/pdfs/Brightmail_Anti-Spam_5.5_datasheet.pdf>.


Brunker, Mike. FTC mulls bounty system to combat spammers. 30 June 2004. MSNBC. 8 July 2004


Falk, Ed. Tracking the source of email spam. 25 June 2004
<http://www.rahul.net/falk/mailtrack.html>.
A Systems Analysis of the Spam Problem
<http://download.microsoft.com/download/d/a/2/da2821f5-6acb-4058-8974-5a3c7d187794/senderid.pdf>.


<http://www.oag.state.va.us/Protecting/SPAM_TECHNOLOGY/spam.htm>.


Wikipedia. Bayesian probability. 30 June 2004. 22 July 2004

Wikipedia. Domain Name System. 20 July 2004. 20 July 2004

Wikipedia. IP Address. 23 July 2004. 15 July 2004

Wikipedia. Open proxy. 1 July 2004. 18 July 2004


Wikipedia. Public-key cryptography. 21 July 2004. 21 July 2004

Windigo. Spam: A Look at the numbers. 1996. spam.abuse.net. 22 June 2004


Yahoo! Inc. Computers and Internet > Communications and Networking > Email > Protocols > SMTP (Simple Mail Transfer Protocol) > RFCs. 2004. 13 July 2004

Yahoo! Inc. Delivery Problems. 2004. 1 July 2004

DISCLAIMER OF QUALITY

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available. If you are dissatisfied with this product and find it unusable, please contact Document Services as soon as possible.

Thank you.

Some pages in the original document contain pictures, graphics, or text that is illegible.