1 Project Description

1.1 Introduction

Anyone who has ever had an email account has received messages containing unwanted advertisements (commonly called spam or Unsolicited Commercial Email [UCE]). The problem of spam has spawned fierce debate, considerable software development (both for sending it and for filtering it), and plenty of legislation. Despite all this, there are lots of marketers who continue to send spam and millions of email users who want to get it out of their in-box.

Spam has become a big problem because it allows mass marketers to shift the costs of delivering their advertisements to the recipient in a way that is not possible with traditional media. When an advertiser sends you junk mail or calls you on the phone it costs them a small amount of money each time they do so. With spam the cost to the sender is miniscule, with most of the cost is absorbed by the receiver and their Internet Service Provider (ISP). This results in marketers being tempted to send out their advertisements to hundreds of thousands of people knowing that even a 0.1% response rate will make them a profit. It also opens the door for advertisements of products like pornography or illegal cable descramblers which many people don't wish to receive.

The problem is just as bad for the ISPs who process the mail. It's estimated that as much as 75% of email traffic is spam (see <http://www.emarketer.com/estats/020199_email.html>) which means a significant amount of an ISP's bandwidth is consumed by spam. The situation is particularly bad here in Hawaii and around the Pacific Rim where Internet bandwidth costs many times what it does on the US mainland.

1.2 Existing Solutions Are Not Effective

The problem of spam has not been solved to date because it is a complicated issue with economic, social, and political facets. Existing attempts to solve the problem include:

- Legislation. Bills have been proposed (and sometimes passed) at the state and federal level. Some bills seek to legitimize spam leading to fears that this might 'open the floodgates'. Other bills seek to criminalize it which some feel would be a bad precedent for free speech.

- Email client filters. These are usually limited to clumsy user-generated rules which attempt to catch spam and delete it. Keeping the rules up to date is tedious and error prone, and each program has a different way of setting up filters.

- ISP server filters. These filters usually work based on the source of the message so they are limited to catching spam runs which have already been detected and reported by others. In addition, they typically reject messages wholesale which takes the decision out of the hands of the recipient.

1.3 Adaptive Personalized Spam Filters

I am proposing a hybrid technical solution to the problem of spam: filtering on both the user and ISP side. The filtering program on the user side is predicated on differences between a spam message and a normal message. A series of tests will be performed on each message in a users incoming mailbox . Each of these tests will attempt to determine if the message is spam. Some of the obvious tests that can be used are: is this message actually addressed to the user (spam frequently is sent with the digital equivalent of "Current Resident", is this message from a previous correspondent, does the message contain telltale phrases

like "MAKE MONEY FAST". More subtle tests will look for things like: excessive CAPITALIZATION, excessive use of exclamation points, lists of addresses for pyramid schemes. However, unlike most filtering systems each test will produce a probability instead of a yes or no result. The accuracy of the tests will be determined by feedback from the user after examining the actual mail. The probabilities generated by each test will be combined into a 'spam probability' for each message. The spam probability can be used in several ways: users can sort their email so that the messages that are probably spam are shown last, or users can set a threshold probability above which messages are automatically deleted or filed away.

This system has advantages over existing client filters. The system doesn't have to make binary decisions about spam which allows users to set their own tolerance for spam. It also adapts to the mail patterns of the individual user without requiring the user to manually figure out what rules to input. However, it leaves the final decision in the hands of the user.

On the server side, the system will examine each message as it is received and assigns a spam probability. This would allow the ISP to completely reject some messages based on initial information based on the recipient's stored profile, thus saving bandwidth.

This system, if widely implemented, has the potential to not only greatly reduce spam received, but it can contribute to reduced **production** of spam. When spam is filtered out of most users mailboxes, spam will cease to be a useful marketing tool and spammers will stop spamming.

1.4   Business Model

I plan to market the client software and server software differently. I will give away the client filtering software, allowing users to start getting rid of spam immediately. The client software will not require their ISP to install any software (which some ISPs are reluctant to do and even when they want to it can take months). However, I will sell the server software to ISPs as they are more likely to be willing to pay since they can simply account for it as a cost of doing business. The free client software will also act as an excellent advertisement for the server software: users will download the free software and encourage their ISP to purchase the server side. This is similar to the successful "free client/costly server" business model used by Netscape, Sun, Microsoft and others.

1.5   Project Plan

I will be working on this project after graduating in Spring 2000. I am familiar with the tools required to write the software so I expect an initial version of client software can be completed in a few months. At that point I will release the client software to the world so I can get actual user feedback. The server software will be more complicated because ISPs use a wide variety of mail server software so work on the server software will begin after the client is released.

1.5.1   Deliverables

By the end of the grant award period, I expect to have a working version of the client software in general public release. Depending on the demand from users, I may also have a version of server software available. I will also have a report describing the progress I have made on the project and relating the lessons learned in the process.

2 Project Budget

The project budget consists of two items: a desktop system which will be used to do the actual development of the software, and a server which will be used to test the server software. A development system is necessary because the development will be done after I have graduated thus preventing me from using UH facilities. The separate server unit will be necessary because the server will have to be located somewhere with permanent Internet access (possibly colocated with an ISP) in order to actually test the filtering.

Item	Price	Notes
Dell Dimension XPS B desktop PC	$3,483.00	Or comparable system
VA Linux StartX-SP server	$1,440.00	Or comparable system
Order Total	$4,923.00	

3 Essay Question

As an Aspect Technology Fund grant recipient, how would you contribute to the field of technology and promote the spirit of entrepreneurship?

As a grant recipient I will be helping the growth of high-tech industry here in Hawaii. In the last few months it seems like all politicians and the media can do is talk about high-tech and what it can do for Hawaii. In his 2000 "State of the State" address the governor highlighted technology as an area where the state needs to continue to expand. But it takes more than words to build a business. It takes people actually willing to take risks and make a commitment to hard work.

I have already been involved in one successful high-tech startup here in Hawaii: LavaNet. I left the MS program in ICS to help start LavaNet in 1994. Through a lot of hard work, in 1997 LavaNet had grown to employ 23 people and serve 6,500 customers. At that time I decided to come back to UH to finish my degree.

One of the very real problems with the high-tech industry here in Hawaii is the migration of talented people to the mainland. The booming Silicon Valley economy makes it difficult to convince technical people to stay in Hawaii. My experiences at LavaNet I have enabled me to evangelize the concept of entrepreneurship with first hand experience. Many students here at UH have not seriously considered being part of a startup company here in Hawaii as a career path which is unfortunate. This grant would form the foundation of a software business based here in Hawaii. Beyond the initial implementation, developers will be needed to enhance existing spam tests and add new ones. The server product will require a skilled technical support staff in order to help ISPs install the software. There is a lot of potential for this kind of enterprise because for information technology, Hawaii's geographical isolation is irrelevant. Also, information technology has virtually no environmental impact compared to other industries, which makes it sustainable.

I'm also pleased to hear that there is more venture capital available for high-technology now than ever before. Most of the credit for this must go to the incredible success of the Digital Island IPO. In my opinion the best way to promote the spirit of entrepreneurship is to lead by example. Each successful technology company further demonstrates that it is possible (and profitable) to start up in Hawaii which encourages others to take the entrepreneurial leap. And those that payoff big for their investors will provide the seed funds for more companies (as Digital Island has done).

In conclusion, I have the experience and the ability to take an idea and turn it into a company. Every high tech company that starts up and succeeds in Hawaii provides another beacon for future entrepreneurs.