RSS Security in Ethernet Network Protocol

BY

ADEBIAYE FAPETU
MASTERS OF SCIENCE IN INFORMATION SECURITY
2007

ABSTRACT

Distributing information has come a long way from the traditional mailing system to emails, SMS (Short Message Service) and the likes. RSS (Really Simple Syndication) is another form of distributing information through the web.

With the use of an RSS, the user can read reader (or aggregator) RSS feeds from different websites.

These feeds contain links to information that the reader might be interested in reading. This study covers the designing of a security framework for RSS because presently, the use of RSS has no security from possible intrusion by malwares.

The security framework, which will be implemented on a website, will gather and filter RSS feeds, and then, redistributes these feeds to the designated reader.

XML (Extended Mark-Up Language), on which RSS is based, is the main programming language used in this study. URL filtering will be used primarily, along with other filtering concept to achieve a multi-layered filtering system.

Objective of this project

The objective is to present how RSS works in distributing information through the web and how it could be relevant for Healthcare communication. The security concepts, importance, threats and vulnerabilities in RSS will be highlighted.

Secure RSS Syndication

On July 13, 2005 there was an inquisitive question by a user named Joe Gregorio and I quote

“I have a problem. It's actually a pretty common problem. I have data that I want to syndicate to myself, but I don't want you to see it. It's private. Now this could be my credit card balance or internal bug reports for the day job. Either way, I want the information in a form suitable for syndication but not available to everyone”.

What could be the solution to this question?

There is a solution. I could password-protect my feed. But that causes a problem, because my aggregator would then need to know my password. Now my aggregator of choice is Bloglines, and I'm sure they're nice folks, but I really don't want to give them my password. One security breach and my whopping credit card debt are splattered across the Web. There are problems and there are needs for absolute secured RSS syndication. This is the technology of new era, which’s has to be protected.

TABLE OF CONTENTS

Table of figures

List of tables

INTRODUCTION

The Birth of RSS

With the growth in a number of websites abounded, there were just too many websites that carried out information of interest for every one of us. There was a 'need' to monitor the changes that were happening in various websites of interest to individuals. If only there could be some means by which the changes happening in every web site could be monitored, then the new happenings across the Internet world could be reported back to a 'news room'.

In order to make this a possibility, the screen scrapping software was written, which could bring about or monitor changes that are happening on any specific website. Some of the most common requirements for this were the responses that one might get for his blog entries. In Figure 1 below: We could infer that once the blogs are entered, every author of the article would like to know when some one would give a comment or when some one would make a remark related to his blog entry. The second major need was in the news sector. When there was capital news occurring on a news portal, many people wanted a news bite. This could also be some news connected with stock prices of individual companies. All these meant that the information available at one point on the network needed to be monitored and shared or syndicated across the entire Internet.

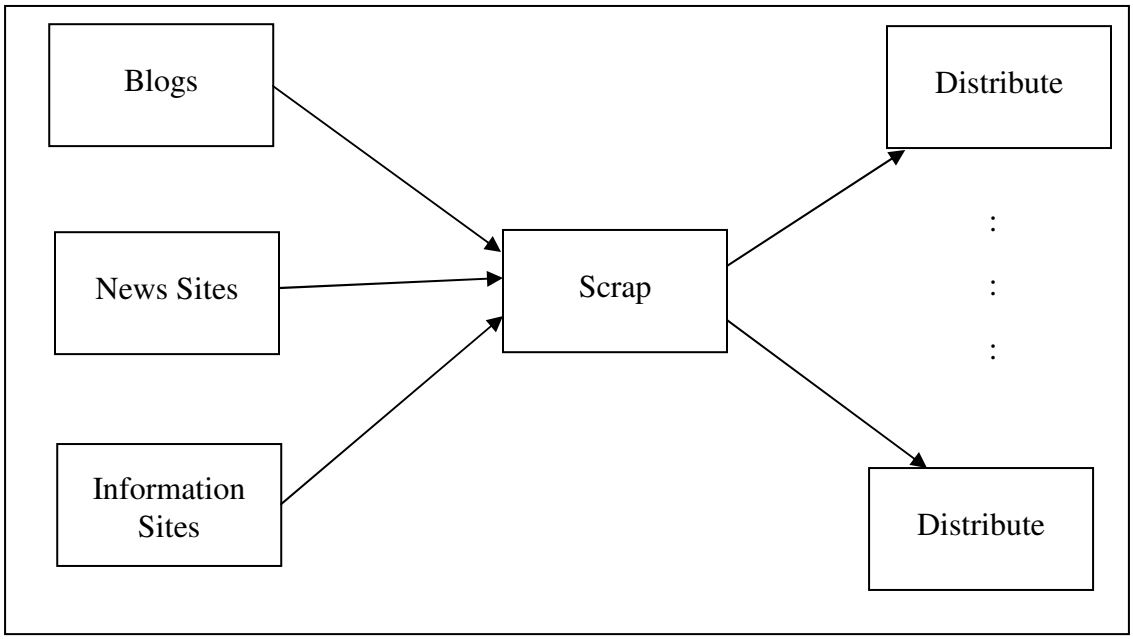


Figure 1: RSS Sources and their distribution.

In order to achieve this, a preliminary form of syndication services was started. Once this caught up, standards were evolved and the syndication itself became a success. This technology was called RSS. As in the case of the technology for RSS, there are multiple definitions for the initials 'RSS'. RSS is supposed to stand for:

a. Really Simple Syndications
b. Rich Site Summary
c. RDF Site Summary

Whatever be the definition, the core of the RSS is similar. It is an XML file that informs of the changes that are happening on the web site.

It is called by different names, viz., XML feed, RSS Feed, Web feed, RSS stream or RSS channel. Whatever be the name assigned to it, it all implies the same and produces a similar output that produces almost the same effect. The output of the RSS feeds generally, have the changes that has happened on the site since the last time the feed checked it.

Objectives of this research thesis

This research aims at conducting a detailed study on the security issues that abound RSS feeds. The nature of security risks that are faced and the security vulnerabilities within the RSS are also analyzed. A detailed literary review is conducted and the required information is culled for further analysis. The methodology for collection of information and analysis is first presented. On compilation of all relevant information, it is analyzed and compared with the existing literature to verify whether there is a possibility of taking care of these infringes and if so, how could the be done. These are marked out in recommendations and conclusions.

The major hypothesis of this research is that under the current conditions and specifications of RSS 2.0, it is highly probable that the security of the RSS can be breached. Spammers can still send in data that is not solicited by the receiver using RSS. Similarly, RSS can also be loaded with unwanted viruses and Trojan horses.

This research aims at proving or disproving this hypothesis. However, in the course of the survey and the later research and analysis of the data, it was found that the hypothesis is held valid.

Methodology

Literary Review

A literary review of available literature on the objective is undertaken. Papers and information available in various sources are first used to understand and present the basic technology behind RSS. In Figure 2 below: Literature pertaining to the break in security and the possible gaps that could lead to security lapses are also studied. Contemporary papers from various sources and books related to the specified topic are gathered and the information so collected is presented in the paper. This information is then used for further analysis of the objective of this thesis.

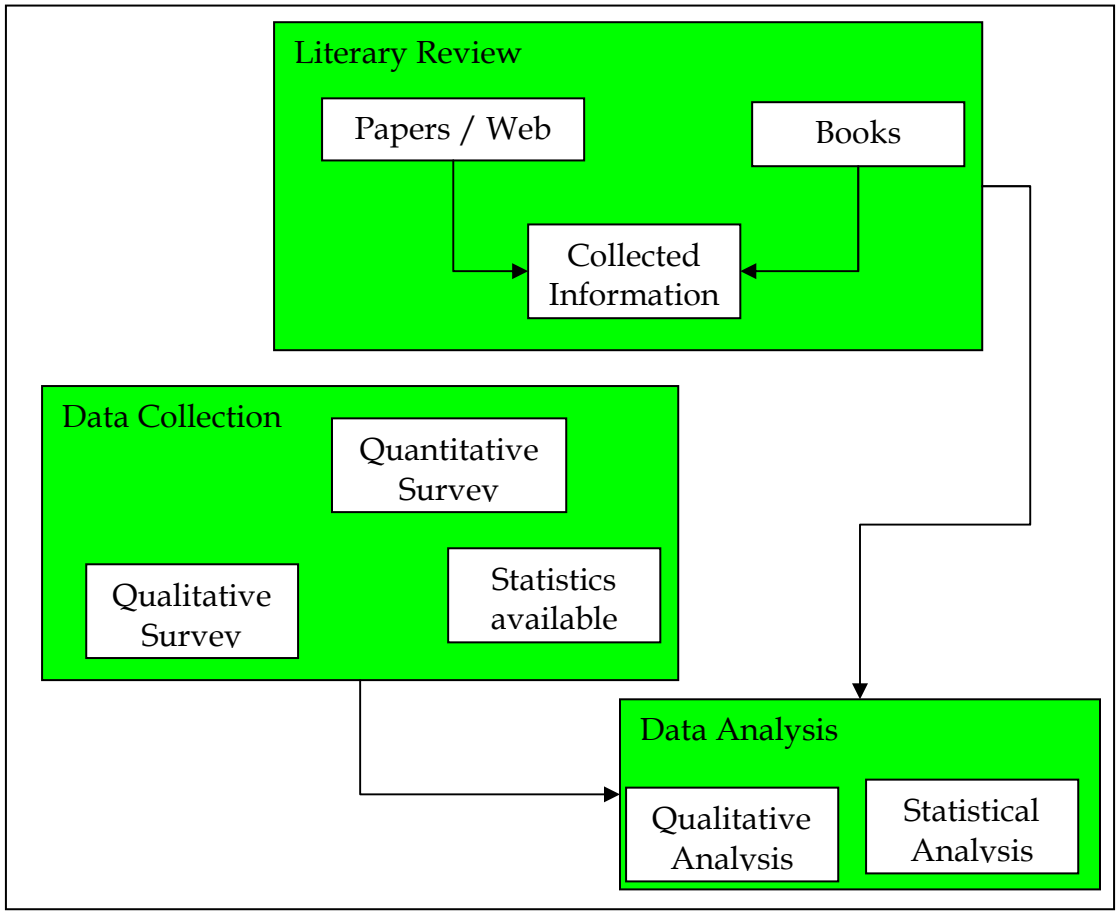


Figure 2: Methodology

Data Collection

Data collection is done in two stages. One, information from the varied statistics that is available in various literatures and statistical databases, and two, from the survey that was conducted during the course of this study. Data on the incidence of security lapses and issues pertaining to RSS feeds are available from various databases and statistical information. This is gathered and will form the generic data for the issue at hand. The second set of information was gathered when the survey was conducted. This was done using a questionnaire that had both qualitative and quantitative questions. Open questions gave way to qualitative responses from the respondents. While the quantitative questions, provided quantitative answers that were subsequently used along with the statistical information for analysis.

A questionnaire was framed for this purpose. The questionnaire employed is provided in the annexure 1.

Respondents were selected on a random sampling basis. No segregation of the respondents was made. The respondents were all making use of RSS and those who are not making use of RSS were also included in order to collect relevant reasons as to why they are not using the same. This information was also used to analyze subsequently the reasons behind the security failures and the extent of security failures that has been happening in the RSS feeds.

Analysis

Information and data collected using the questionnaires from the respondents were tabulated. This was then analyzed and the general formats of the responses have been shown in the figures during analysis.

Based on the outcome of this analysis and on the results of the literary study from which substantial statistical data has been assimilated, further analysis is carried out. This has helped in identifying the extent to which security breach is happening in RSS feeds. Security issues other than unwanted information and the methods normally adopted by the spammers are also studied based on the outcome of this information. This analysis resulted in drawing specific conclusions in support of the hypothesis of the research objective. Adequate statistical data is produced to amply prove that the research objective and the hypothesis have been met.

THE CONCEPT OF RSS IN INFORMATION DISTRIBUTION

The World Wide Web (www) or simply put, the web, is known as an information-browsing application, which generally allows users to locate and access information stored on a remote server and also, to follow references from one server to similar or related information stored on another server. It is important to note that the web is an immensely comprehensive operational network of networks. As such, the web presents a number of unique challenges in the context of information distribution and security.

If we study the conventional way the news (information) has been transmitted between media or organizations, we may identify four important elements:

The Protocol – the rules that govern the packaging and transmission of data between parties.
The Envelope – An information segment, consisting of header and body. We might refer to this as the information item itself.
Header – A component of an envelope that expresses the envelope's metadata
Content – The actual content of an envelope.

With this picture in mind, what's required in creating content for paper, web and other media is now more demanding on the news production system. This becomes overwhelmingly demanding on the transmission and storage of news content and correlations of providing precision information. The web, however, has boosted the requirement for automatic handling of news content, which needs to support more granularity than straight text. It now provides a cleaner, more reusable approach, thereby, bringing about what we call *News Syndication.*

News syndication could be explained as transmission and storage of news content, more specifically, the granularity, structure, and precision of information **[1].** It allows for simultaneous creation of content for paper, web and archive destinations as part of the same news production system. Typically, human intervention is employed at many stages in the production process of print publications. The idea behind modern news syndication is to eliminate this manual, customized processing and streamline the production of news content.

The web has become a tool for news syndication by employing what is generally known as *The Standard Generalized Markup Language* (*SGML*). This is a metalanguage in which one can define markup languages for documents. A *markup language* combines text and extra information about the particular text. This extra information, about the text's structure or presentation, is expressed using *markup*, which is intermingled with the primary text**.** The best-known markup language in modern use is known as HTML (Hyper Text Markup Language), which is one of the foundations of the World Wide Web.

Due to SGML complexity and the need to have a lightweight approach and to make it simple for general-purpose applications, such as Semantic Web, that is, to create a universal medium for information exchange by putting documents with computer-processable meaning (semantics) on the World Wide Web, the *XML profile was designed.*

XML is a profile, that is a specific subset of SGML, designed to be simpler to parse and process than full SGML, and to have more lightweight internationalization.

XML means *Extensible Markup Language (XML*). It is a W3C-recommended general-purpose markup language that supports a wide variety of applications. W3C is the *World Wide Web Consortium (W3C*).**[2]** It is the main international standards organization for the World Wide Web. XML languages or 'dialects' are easy to design and to process. XML is also designed to be reasonably human-legible, and to this end, terseness was not considered essential in its structure.

The primary purpose of XML is to facilitate the sharing of data across different information systems, particularly systems connected via the Internet. With this in mind, it became obvious that a simple format is needed to facilitate news syndication on the web.

An initial idea of "Rich Site Summary", RDF (Resource Description Framework) site summary came up, and after much evaluation, "Really Simple Syndication" (RSS) standard for a generic specification of data formats was finally adopted. As RSS gains momentum, security fears loom large. As publishers are quickly finding innovative uses for RSS feeds, hackers are taking notice. The power and extendibility of RSS in its simplest form could also be its greatest disadvantage in information security.

The constant requirements and need of users, for news syndication made implementation of web feeds inevitable. Web feeds is a data format introduced to serve users' frequently updated content. **[3]** There are the Content distributors who manage this news *syndication of* web feeds making users to *subscribe* to it. This data format of web feeds is of different modifications depending on the Content Distributor and what is subscribed for. The general idea is to make a collection of web feeds accessible at one spot, called "Aggregation". Amongst this family of web feeds data format is, what is generally known as, RSS **–** *Really Simple Syndication.*

Characteristics of Web feeds

The implementation of web feeds will require a content provider, who will publish a feed link on their site which allows the end users to register with an *aggregator* program, or a feed reader or a news reader having control of the program. A typical web feed logo could also be called an RSS logo as seen in Figure 1: The mode of doing this is to bring the link from the web browers to the aggregator. In doing so, the aggregator determines when to download new content at periodic intervals. Contents delivered by a web feed are typically HTML (webpage content) or links to webpages in summaries' form rather than in full content, especially on the updates.

Figure 3: A typical web feed logo

Uses of Web feeds

In order to understand how web feeds are processed, it is absolutely important to understand the work of the "Aggregator". Figure 3 shows a typical web feed logo. An *aggregator* also known as *news aggregator or feed reader* or even called *search aggregator***,** because of its role of using web feeds to retrieve syndicated web content In figure 4: Aggregators reduce the time and effort needed to regularly check websites for updates. Once subscribed to a feed, an aggregator is able to check for new content at user-determined intervals and retrieve the update. "The content is sometimes described as being "pulled" to the subscriber, as opposed to "pushed" with email or IM. Unlike recipients of some "pushed" information, the aggregator user can easily unsubscribe from a feed."**[4]**

These aggregators suck up news feeds from various locations and sites and distribute it to interested users. This way, the interested users get what they need, and content distributors now spread it to users.

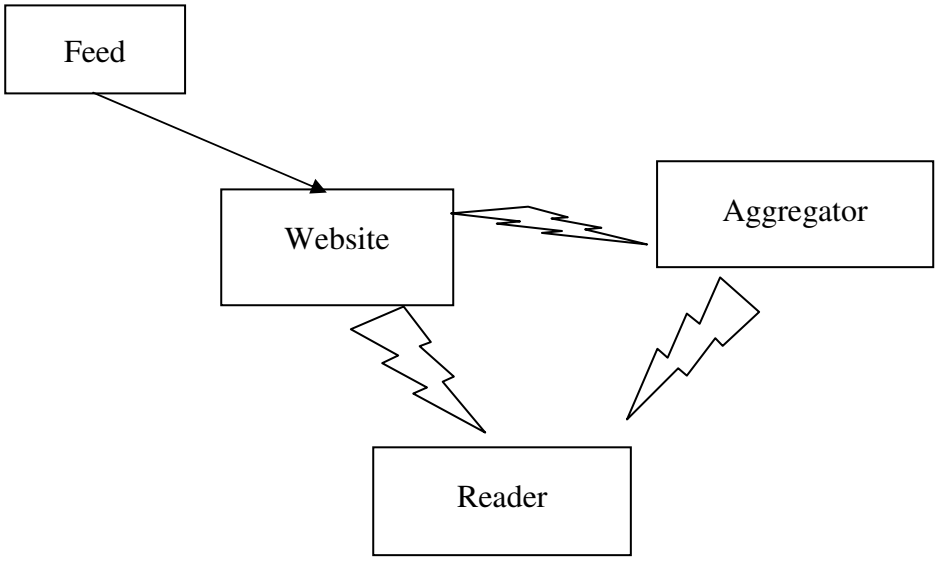


Figure 4: Basic structure of a web feeds

History of RSS

RSS took its birth when the technology for RDF, the Resource Description Framework, was created in 1997. The RDF was a forerunner to the basic requirements of RSS. The history could be traced as indicated in the table 1 below:

Table 1: Evolution of RSS, by permission - Creative Commons Attribution/Share Alike license [5]

1	1997	Scripting News format was formulated. RDF was defined by RV Guha.
2	1999	RSS 0.90 version was first designed by Netscape, for use in their website my.netscape.com. This also supported the scripting news format that was originally propounded. However, the major change in the RSS used by Netscape and that in the scripting news was that Netscape also made use of the RDF for providing the headers in addition to the standard XML that was adopted.
3	1999	This year also saw the release of scripting news 2.0b1 version, which was a landmark release in the history of RSS.
4	1999	Late 1999, RSS 0.91 was designed by Netscape, specifications was written by Dan Libby, includes most of the features that scripting news 2.0b1 had. One more important turn was that there was no RDF header this time.
5	1999	The maker of the scripting news, UserLand also adopted the new standards brought in by Netscape. Instead of two competing technologies and standards, there was now, one single standard and every one could adapt to this.
6	1999	By the end of 1999, Netscape thought RSS is not going to catch up and dismantles the team at its research unit. This brings the RSS development at Netscape to screeching halt.
7	2000	UserLand continues to live with RSS 0.91 and continues their work on it. They release the new specifications for RSS 0.91
8	2000	RSS 1.0 gets published as a proposal developed and brought in by a private group of people at O'Reilly. This used RDF and namespaces. RSS 1.0 was drastically different from RSS 0.91 that was propounded by Netscape and UserLand.
9	2000	UserLand releases 0.92 versions of RSS and also discusses of releasing 0.93 versions but does not release it.
10	2002	RSS 0.92 was used by MetaWeblog to make an API that became an instant success with XML-RPC. This also rekindled interest in RSS to a great extent.
11	2003	RSS 2.0 was a retrofitted with 0.92 version, with additional formats and design elements brought in. This was released on July 15 2003, through Harvard under a Creative Commons license. From then on, the control to monitor and to augment specifications of RSS rested with RSS Advisory Board formed by Harvard University.

With these changes happening in the specifications of RSS, the changes in the RDF also had a significant impact on the RSS operations methodology. A number of competing projects were done in Europe and elsewhere in the form of Dublin Core framework and others. But more need to be proved to ensure that any one of them could compete with RSS effectively.

Really Simple Syndication(RSS) is, therefore, used to publish frequently updated digital content, such as blogs, news feeds or podcasts**.** A **blog** is generally a user generated website where entries are made in journalistic style and displayed in a reverse chronological order – like having a personal online diaries, while a podcast is usually a media file that is distributed by subscription (paid or unpaid) over the Internet using syndication feeds, for playback on mobile devices and personal computers. Like 'radio', it can mean both the content and the method of syndication.**[6]**

RSS starts with the feeds getting created. The feed works in most cases as an update with a topic and a short description. This entry is made in the feed file that is maintained in the site. The feed thus created and updated continuously is monitored or read by two programs outside the website. One is the aggregator and the other is the reader. While reader is more individualized and works normally in conjunction with programs like Outlook, Aggregators are also redistributors. They collect the feeds from the website and feed to other readers / aggregators as seen in Figure 2.

The Feed is an XML file that is defined under the RSS standards. **[7]** RSS 2.0 is the general purpose RSS feed standard that is adopted by the RSS Advisory Board, which works in conjunction with Harvard University, which decides the basic structure of the XML file that works as a feed. Figure 5: Shows an example of RSS standard distribution. All RSS feeds are XML 1.0 files as specified by the World Wide Web Consortium (W3C). This structure is open and is fed by the specified website or is extracted out of the specified website. Therefore, it was assumed that the technology does not permit virus or security infringes easily.

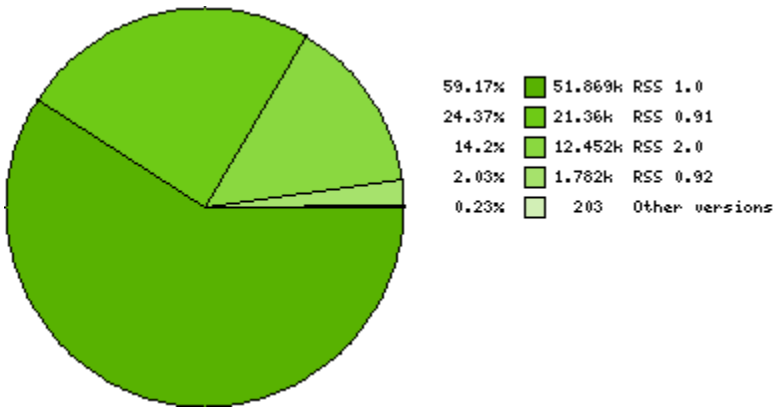


Figure 5: RSS standards distribution (culled from www.rssfeeds.com 2006) [8]

However, it has been found that though the technology is clear, it is quite possible for hackers to use feeds to hack or spread viruses over the Internet. In figure 3, the statistics

show RSS standard distribution. The myth of the RSS being a secure distribution mode for information has been lost.

Purpose of RSS

To keep the competitive edge and increase online exposure we need to use RSS feeds to distribute content. Since pod-casting is distributing audio content using RSS, making audio content available using RSS, pod casters give listeners more control over what they listen to and when, and can easily be syndicated Table 2 below highlights the purpose of RSS.

Table 2: Simple breakdown of RSS process

Instant Information	Information updated in real-time
Single Source	All information aggregated at a single location
Rapid Scanning	Feed readers highlight unread headlines
Categorizing	Information can be categorized by theme

With the use of an RSS reader, or aggregator the user can read RSS feeds from different websites. These feeds contain links to information that the reader can read. RSS readers, news aggregators, or pod-catchers automatically download the information contained within, regardless of its file type or source and there is the inherent risk of an infected file being distributed or downloaded. In addition to displaying news on other sites and headline viewers, RSS data can flow into other products and services like PDAs, cell phones, email ticklers, and voice updates. Distribution of malware, viruses and spy applications may strike the users while they "choose" the content that they receive or download. It is important to include filtering, screening or authentication capabilities in RSS feeds.

Why RSS Feeds?

RSS Feeds were brought in as a clever way of getting around all those people who are trying to send in Spam mails. In order to avoid this, a separate server was used to ensure that only that information as requested by the respective user, is delivered to him. It is estimated that nearly 63% of the consumer goods marketing people will be using RSS to market their products. 65% of the media and communications marketers will also be doing the same thing to effectively reach their clients. Nearly 40% of all retail marketers, financial services marketers and tech equipment marketers (Advertising Age, 2007) will continue to use the system in their effort to grab the market share. **[9]**

RSS feeds were directly made and fed by those who are managing a website and the feed is normally a registrar of changes in the website. These changes are then picked up by the RSS feeder and passed on to the server for the RSS. Therefore, it was felt that the source of information is authorized and there will be a decrease in the Spam transaction over the network. First, the Spam continues to exist though the effectiveness of Spam has come down drastically, especially with users protecting their mails with several anti-virus

software programs. Secondly, the RSS has not yet become the method for communicating with the customers. RSS, in most of the cases, are auto generated when the changes are made on the site of interest.

When a feed is created, it is common to find that it first gets validated using a feed validator. A number of feed validators exist to check and let the user know whether the feed is a valid RSS syndicated feed or not. Most of the validators work in conjunction with multiple versions of RSS, typically, RSS 0.90, 0.91, 0.92, 0.93, 0.94, 1.0 and 2.0. Standard validators are available on the net.

References:

[1] Dumbill, Edd "XML in News Syndication" July 17th, 2000.

http://webservices.xml.com/pub/a/ws/2000/07/17/syndication/newsindustry.html

[2] Pilgrim, Mark "What is RSS" December 18, 2002

http://www.xml.com/pub/a/2002/12/18/dive-into-xml.html

[3] http://www.silverpop.com/rssdirect/index.html

[4] Wikipedia "RSS (file format)" (accessed on 31 December 2006)

http://en.wikipedia.org/wiki/RSS_(file_format)

[5] http://www.rss-specifications.com/history-rss.htm

[6] http://en.wikipedia.org/wiki/Blog

[7] RSS Advisory Board, 2005, RSS 2.0 Specifications, available at:
http://www.rssboard.org/rss-specification

[8] RSSfeeds.com, 2006, RSS Statistics available at:
http://www.rssfeeds.com/rss_stats.php

[9] Advertising Age, 2007, Interactive Marketing and Media - Fact Pack 2006, A supplement to Advertising Age, available at:
http://adage.com/images/random/Interactivefactpack06.pdf

RSS Technology

RSS FEEDS DISTRIBUTION

An RSS feed is a list of items or entries. Every entry links to the relevant information in the web site. RSS looks at the information in a web site as data and uses metadata to represent this data. The metadata may consist of a link to the relevant information, a short description of the data, a title or a headline and others. The content of the data that it has could vary from one feed to another. **[1]**

A typical RSS feed might look like the one below:

```
<Item>
  <Title>Weapons of Mass Destruction Found! </title>
  <Link>http://news.example.com/2006/10/WMD</link>
  <Description>the troops that over ran Iraq, at least
found what the army claimed as the Weapons of Mass
Destruction. General Patrick addressed the press
conference today to release this news to the press.
</description>
</item>
```

The RSS feed might have a heading in addition to the item heading that the feed has **[2]**

```
<? Xml version="1.0"? >
<rss version="2.0">
  <Channel>
    <Title>News Channel</title>
    <Link>http://news. COM/</link>
    <Description>The first news Channel</description>
  <Title>Weapons of Mass Destruction Found! </title>
  <Link>http://news.example.com/2006/10/WMD</link>
  <Description>the troops that over ran Iraq, at least
found what the army claimed as the Weapons of Mass
Destruction. General Patrick addressed the press
conference today to release this news to the press.
</description>
    <Item>
       <Title>News for September the First</title>
       <Link>http://example.com/2002/09/02</link>
    </item>
  </channel>
</rss>
```

- The news channel will become the name or the title of the feed and this might have a list of items, which is then presented through the feed.
- The feeds can provide for the following metadata, in addition to the common ones like link, title and description metadata.
- The Image will now allow the feeds to carry thumbnails image relevant to the feeds.
- *Webmaster and managing Editor:* The person who is responsible for the feed.
- "*lastBuildDate*": Date of last update of the feed
- *Enclosure*: to allow an automatic download of an attachment to the feed
- *Guid*: to identify the item uniquely in the list. The usability of this varies with the aggregator used. Figure 6 also shows a typical RSS feeds distribution as related above.

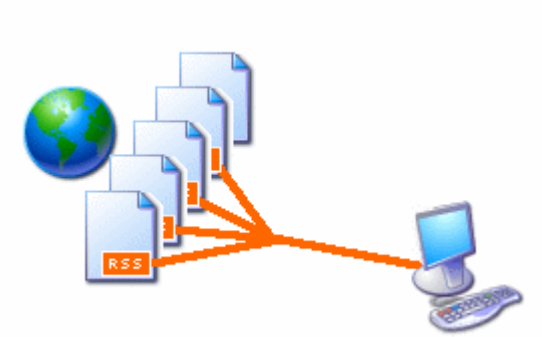


Figure 6: Typical RSS feeds distribution [2b]

All of these metadata would modify the content of the feed and would also provide the required information to the aggregators and feed readers.

Steps in creating RSS Feeds

RSS feed is created by using feed generators that gather input from the website and processes information to identify the change that is happening at frequent intervals. This process is a part of the syndication exercise that RSS does. This enables the Feed Generator to identify the nature of change that is happening on the site and generate appropriate titles and elements for the list entry.

This gets fed to all the subscribers of the feed using the aggregators. It is important that the RSS feed follows the best tips mentioned below **[3]**.

It is important that the feed has distinct entries to identify from where the feed is originating. This has to be done using the guid key in RSS 2.0 so that the aggregators are able to differentiate the feeds from the others.

The title and the description on the link should be appropriate. The title should not be too short. This will not generate enough interest in the article and might not convey much to the readers. On the other hand, the entire article should not go on the description, for that might be a waste of time for the reader and the user might consider it cumbersome.

It is safe to avoid using html code embedded in the feed since they are not represented directly and might change in the way it appears to the user from the way it does during the test. None of the html entities are available in the XML like; & copy. Therefore, in case if it has to be used, it must be redefined in the XML before use.

It is safe to use Unicode structures, typically UTF-8, for character encoding.

Appropriate tags are to be used and description tag should not be used for multiple purposes. For instance, appropriate copyright tag should be used to display copyright notices and description should not be used to show the copyright notice.

In most cases, it has become a common practice to use software tools that generate the feeds when any change occurs on the website. Some of the typical tools available for this purpose include “xpath2rss”, “myRSS”, “ROME” and many others. There are also validating tools that will help in identifying if the RSS feed generated is in line with the standards and whether the aggregators will be in a position to make use of the feed thus generated. **[4]**

In most of the cases, the content creator creates the feeds. In some cases a third party might be interested in creating such feeds whenever a change happens in the specified web site. This process of Scrapping captures the changes in the web site and puts it through the web feed in the needed format. Most of the automatic generators of web feeds work on this technology and provide the feeds to the users. It could also be termed screen scrapping, this would help in generating web feeds from other inactive sites

Selecting the right Format

The RSS feeds can be in any of the formats that are available for every one of the standards. For RSS version 1.0, the successful formats have been the Dublin Core metadata format. This has been adopted wherever there was version 1.0 in use. The version 2.0 was the one explained in the previous section and the structure is laid out by the RSS 2.0 specifications made by UserLand’s Dave Werner.

RSS 2.0 has been adopted by most of the users of syndication. However, while some joined together to produce another format that is usable in place of the RSS 2.0 formats. This became the Atom later on. The Atom and the RSS are the two competing formats that are used extensively by developers to link up details on various sites. **[6]**

Selecting either the Atom or the RSS for feed would depend on the individual's choice. However, it is found that most of the blogs seem to be using Atom while the news feeds seem to be all RSS. Though there is no major difference between them, Atom also supports the essential tags that RSS also supports like the title, description and link. However, there are others that are changed but there are other tags that do the same job in Atom.

Therefore, it would only be a surprise that Atom was selected for a job because it just outperformed the other. But then what needs to be noted is that Atom is really a development over RSS 2.0; therefore, there are specific formats that make it better than RSS 2.0.

The essential difference occurs when Atom insists on a unique identifier for every feed that is sent out from the web feeder. This makes the Atom feed a more identifiable and traceable feed. Secondly, Atom allows Podcasting that makes the work all the more interesting and distribution becomes easier.

The Atom content model makes it easier for the developer to use any type of content in the feed that is produced. This makes Atom powerful to include text or any other type of content that is needed.

All these differences make Atom a developed version of the format that is made up as RSS 2.0. While RSS 2.0 formats are frozen and the specifications will not be further developed or clarified, the Atom Publishing Format is in line with RFC 4287 and is the new IETF news feed format that has been adopted**. [7]**

In addition to this, Atom uses the IRI format (Internet Resource Identifier) for the purpose of identifying the feed and the other resources that are referred to over the feed. IRI is the special form of URI (Universal Resource Identifier) that would allow the identifier to have special Unicode characters. This will help the web names to have native language names over the web.

In most of the cases, the RSS is easily traceable due to various reasons. One, since one of the aggregators, the source of the feed, pick up the data is certainly known. Two, the data, in most cases, is generated by software that works in a predefined format. This makes the data more in line with what is found in the web site. If the web site is audited well, then so is the case with the RSS feed, it is also well audited. In addition to this, the targeted use of the feed is also traceable. This happens, when the feed from the aggregator is used for splogging (Spam Blog) then the same is traceable. The usages of RSS feeds have been increasing exponentially. RSS feeds are viewed on all days though as the graph below in Figure 7; this indicates there is an unknown reason for a specific day to be having the highest hit rate.

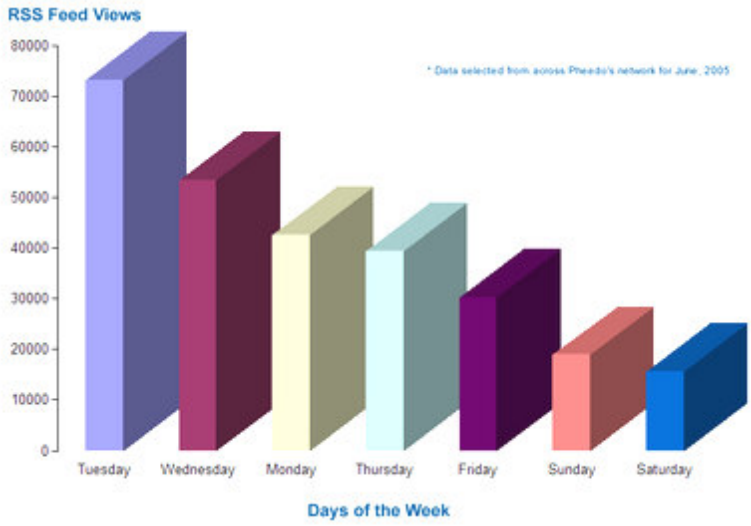


Figure 7: RSS Feed Views and their rise in popularity. [8]

It is important, therefore, to select the right kind of format based on the readership for the feeds. When we look at the security factor in the feeds, it is also found that the maximum reading days are the ones that will also have the maximum-security breaches.

Aggregator

Functions of an aggregator

The feeds from various sources are to be identified and then redistributed to a number of interested subscribers. In figure 8: The Aggregator does this exercise.

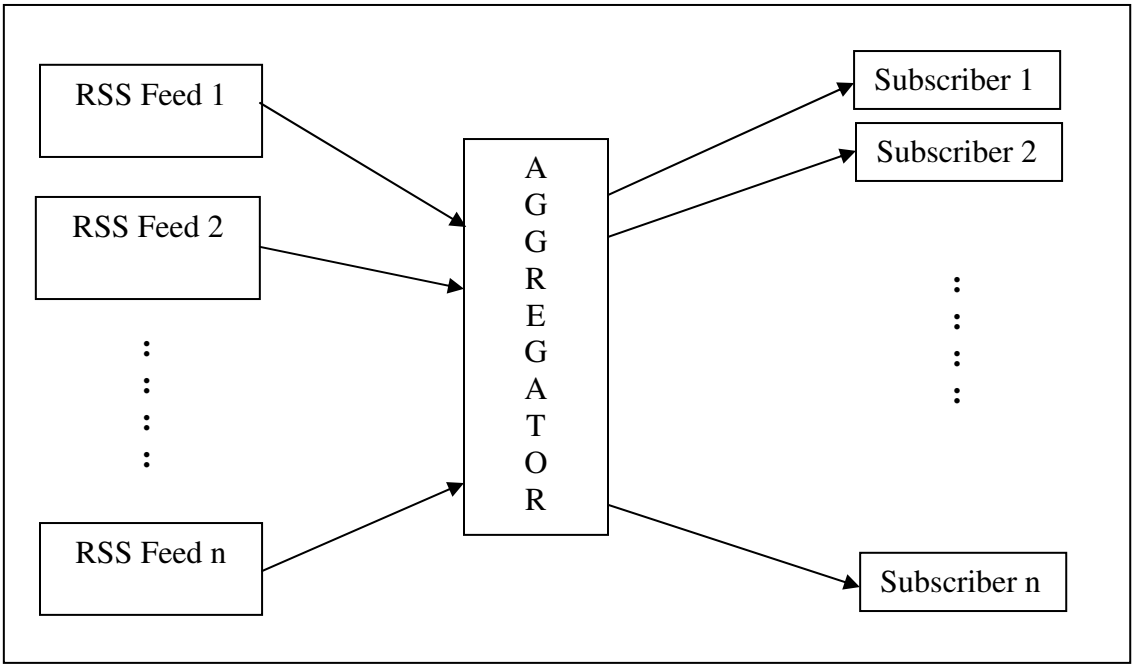


Figure 8: Operation of Aggregators

The technology will make aggregators to collect information from the various RSS feeds sourced from different web sites. These RSS feeds are then identified, classified, segregated and then distributed accordingly to the number of subscribers to this information. Therefore, the action of the aggregators would depend on the format of the RSS feed and the contents of the feed.

Most of the aggregators are classified depending on the type of RSS format that they support. They form the core technology employed at the aggregators. Aggregators pull information from the web site when there is a change in them. Aggregators pull syndicated content at frequent intervals. **[9]** This is normally settable by the user and the aggregators collect this information from the web site at this specified interval.

Aggregators optimize the work by employing two basic methods. One, they provide filtering option to the users. Keyword filtering helps the aggregators to identify and pick up only that information which is relevant to the user and has the key word in it. There is also the exclusion keyword, which helps the aggregators to throw out that information or data which has the key words.

The aggregators achieve the other optimization by using what are called ‘clouds’. Clouds are a form of information source that will have information from various sites stored with them and they would prompt the aggregators or the feed readers whenever there is a change in the specific web site. **[10]** This would save the feed reader from repeatedly cross checking the web site as this would increase the band width usage if a number of aggregators keep checking the web site for changes.

Most often aggregators are contrasted with the push type information distribution achieved using emails or instant messaging systems. In these cases, it is difficult to spot the source and stop such mails, which keep dropping into the mailbox. Whereas, in case of an aggregator, the information is pulled from it; therefore, it is very easy to spot the source of the information and stop the feed.

Implementing an Aggregator for full performance

Aggregators come in two forms: either as the online ones or as the desktop ones. In case of the online aggregators, the information is gathered and distributed either for free or for a charge. The recent changes in the data are shown in the aggregator, which would help the user to pick up the most recent changes from the site rather than the older ones.

RSS Readers

Functional activities of the RSS Readers

The aggregators to be collected by the feed readers at the client end distribute the feeds from the RSS feeds. These take the form of mail boxes in some cases and in others, they are separate feed readers where the user might browse through and read the messages. Individual RSS readers present the final information from the RSS feed and the aggregators to the users.

RSS readers either work in conjunction with the browsers like Explorer and FireFox or they would work as a stand-alone program. With the browsers they look like any other mail client and deliver the messages to the user from the aggregators.

On the stand-alone programs too the messages are delivered as they are done using a mail client. These are accessible as in the case of any other mail client. **[11]**

Every RSS reader would expect to have the feeds selected by the user. This list is the one that is made use of by the RSS reader and is picked up from the indicated web site.

If the feed from a specific reader is not relevant, the user can set it to off and thereby, stop getting any feeds from them. This makes the feeds very flexible and at the control of the user and not at the control of the sender as in the case of the emails.

Installing the RSS Readers

Most of the RSS readers available for usage are all free. Though there are good ones, which might be charged. If the reader is a stand-alone program and can be installed on the system with its current configuration, the same is downloaded and then installed on the computer.

The feeds that are relevant for the user can be subscribed. The sites need to provide RSS feeds using the orange RSS logo. Only these sites can be connected up to the reader.

The feeds and their updates are then received and can be read like any other email. The titles of the changes alone make their appearance in these messages along with a short description.

Sometimes, the files might also carry a downloadable attachment, which is also downloaded by default, the moment the message is clicked. The messages are all XML files on the changes that are occurring on the sites.

References:

[1] Mark Nottingham, 7 Sep 2005, RSS Tutorials for Content Publishers and Webmasters, Creative Commons available at: http://www.mnot.net/rss/tutorial/#Versions

[2] Sharon Housley, 2004, RSS Specifications: Creating an RSS Feed, Note Page Inc., available at: http://www.rss-specifications.com/creating-rss-feeds.htm

[2b] NewsBee RSS reader: NewsBee/RSS Home Page. Retrieved January 14, 2007, from http://www.newsbee.de/branded/ Web site:

[3] Notepage Inc., 2004, Create RSS: Tips for Feeds, available at: http://www.create-rss.com/tips-for-feeds.htm

[4] Paul Miller, 25 Oct 2004, Syndicated Content: it's more than just some file formats, available at: http://www.ariadne.ac.uk/issue35/miller/

[5] Mary Harrsch, Jul/Aug 2003, RSS: The Next Killer App for Education, The Technology Source

[6] Dave Shea, 19 May 2004, What is RSS/XML/Atom/Syndication? Available at: http://www.mezzoblue.com/archives/2004/05/19/what_is_rssx/

[7] Dave Johnson, 2006, RSS and Atom in Action, Manning Publications Co., Chapter 4.

[8] Pheedo, 21 Jul 2005, No.1 – RSS usage Revealed, available at: http://www.pheedo.info/archives/000265.html

[9] Stephens, R T, 17 Nov 2005, Knowledge the Essence of Meta Data: RSS Technology – Evolution, Revolution and Extinction, DM Review, available at: http://www.dmreview.com/article_sub.cfm?articleId=1041761

[10] Wikipedia, 2006, Aggregators, available at: http://en.wikipedia.org/wiki/News_aggregator

[11] Note Page Inc., 2007, RSS news Aggregators, available at: http://www.rss-specifications.com/aggregator-how-to.htm

RSS Security Lapses in Ethernet protocol

Nature and types of security threats

The major issues that throng the Internet world appear to be the following:

1. *The spread of viruses, Trojans and their inheritors*. They spread mostly using the email networks and as the messages are exchanged. It is interesting that the viruses, Trojans, the worms and many other forms of the security breaches, all work to slowing down the systems or break through the existing security in areas where important information for the person concerned is stored. These form the major cause of concern for the Internet community. In figure 9: The nature of virus spreads notices has been remarkably increasing with the increased usage of Internet. Viruses and their related strains continue to be scourge of Internet in the last few years and would continue to be so. **[1]**

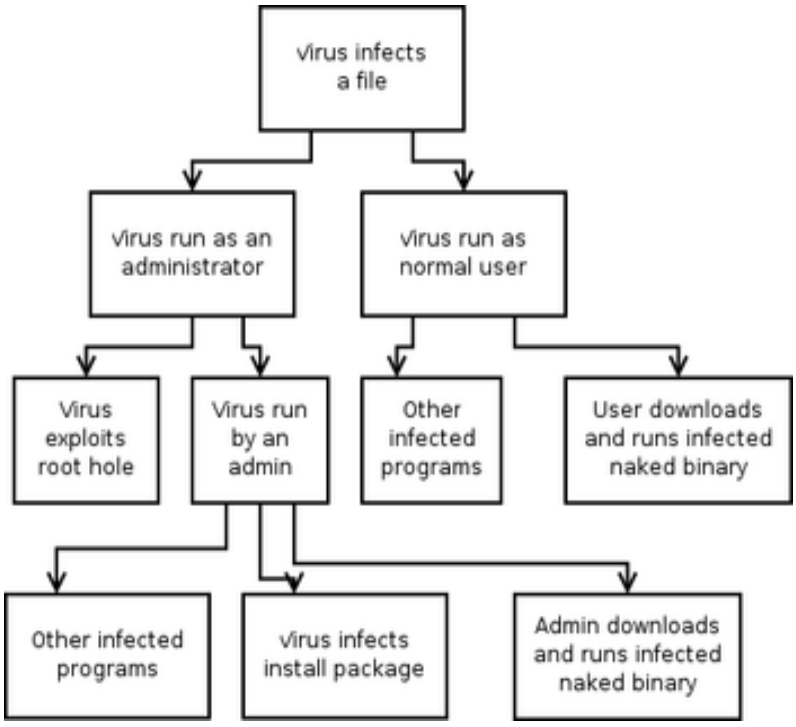


Figure 9: Attack Tree of the Virus [2]

2. *Sniffing*: Sniffing is a technology requirement to monitor the packets that are moving around a specific network. In other cases, where it is not serving for the good of the network community, it is found to affect them on the negative side. In figure 10 below: The information is sniffed out of the network cables when it is being transmitted over long distances. This happens specifically in those cases, where personal information, like

credit card details, bank account details, is moving across the network. This could cause leak of discrete information leading to larger loss to the person concerned. **[3]** Both hardware based sniffers and soft sniffers are available to pick up such data.

Sniffing was developed for standard requirements to monitor movement of packets and information, whereas, in the current context, the information movement is monitored and this could result in the loss of information. The packet sniffing and content sniffing are two types of sniffing that are currently prevalent.

3. *Spoofing*: Spoofing is personification of one as the other. Message, which has come from computer A, appears as if it has come from computer B. This makes the receiver to look at the message and interpret it differently. This can also hurt the way the user approaches user A or user B.
This can be used to obtain information from a user C by A posing as user B. This way, personal and highly confidential information can be obtained. On a larger scale, spoofing becomes what is called the session hijacking when the entire session that the person is logged on with, gets hijacked or moved to a different location. **[4]**

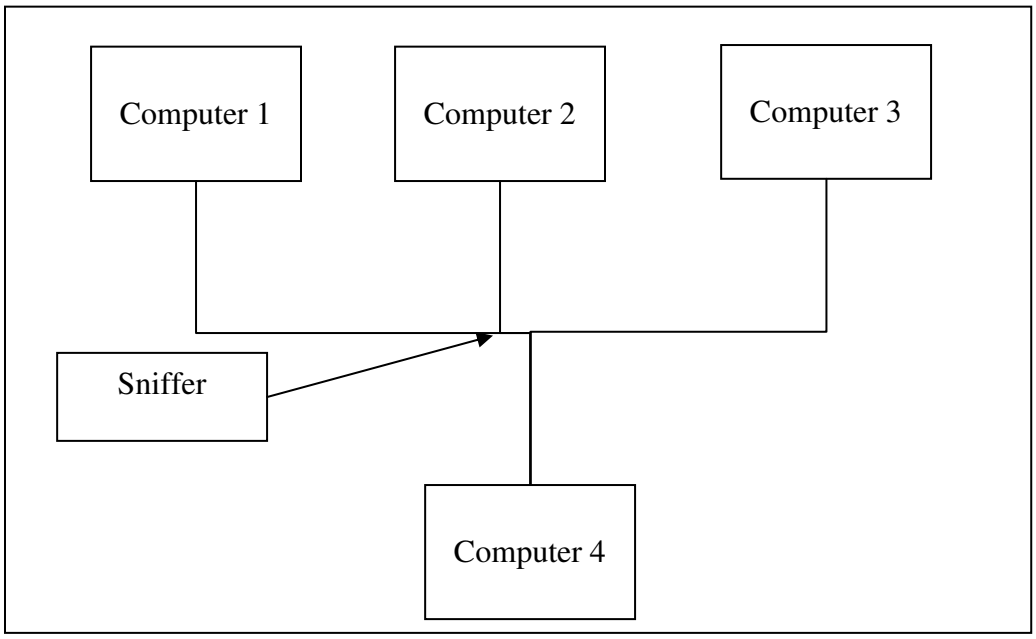


Figure 10: Sniffing on an Ethernet network

Spoofing can be either email spoofing or IP spoofing though both are related to IP spoofing. The emails would appear as if they have originated from person A when actually, person B sent it. This makes emails difficult to trace.

4. *Splogging*: Spam blogging is new to the Internet. However, it has its origins in the same swiftly growing maze of blogs **[5].** They help in increasing the number of blogs in a specified location without adding any value to them. They, like any other virus attacks, also suffocate the web site thereby bringing it to a crashing halt. Secondly and most importantly, splogging is the source of most of the Internet based plagiarism. The screen

scrapping and web site scrapping algorithms collect data from the web sites and provide them to interested people without the consent of the originators. In more than one way, RSS is one of the main culprits to such plagiarism.

5. *Intrusion*: Intrusion of unsolicited visitors on to the private networks is also a major cause for security breaches. Intrusion deductions systems try to identify when there is an attempt at breaking the password or trying to gain entry by letting themselves into another network. Most of the hackers use the password and user identification techniques that will help in authenticating the entry of a person into a private network. This will increase the probability of the person going undetected in a private network. Hard and soft intrusion detection systems are available to identify and stop such intrusions. Identity thefts in United States alone have done considerable damage and caused risk to numerous net users. USA Today reports that from as few as 130 reported security breaches alone, more than 55 million Americans ran the risk of potential identity theft in 2005-06. **[6]**

6. *Phishing*: This helps in identifying the usernames and passwords of various users by employing or combining the spoofing technique, as stated earlier, and the trusting nature of the user. This is a part of the identify theft that is perhaps done on a very large scale. In figure 11 below: A very normally looking email scam that will be the start of the identify theft if the user is not watchful. **[7]** The user is guided to a new site thinking that it is his normal safe site and his username and password in the safe site is obtained. This will increase the probability of intrusion into the safe network as well as in session hijacking. More and more cyber crimes get committed by using this technique.

Figure 11: Typical email used for Phishing [8]

The technologies employed in these techniques were unique and wariness of the user was advocated. However, better devices were made to stop such viruses and other intrusions into the private networks and secure networks. With the onslaught of newer techniques,

no network is secure and no computer is without the risk of running under the attack. With every PC becoming a network since it gets connected to the Internet and to the numerous computers that make up the cyberspace. Every new method adopted to stop the spread of viruses or security risk is soon bent. '*Lest one good custom should corrupt the world*!'

RSS Information Security Bridge

RSS was seen as a cure for most of the security lapses that were inherent in the emailing systems. The emailing systems were pushing the information across to the users; the RSS was pulling information from the web sites and by the users. Table 3 below; shows security and RSS solutions to this security bridge. This made sure that the users were clear as from where and where not to pick up the information needed. Spamming was expected to come down to naught once RSS become popular. But this was not the case because of the major lapses that were confronted by the RSS. The security lapse in RSS is identified as the major cause of worry.

Table 3: Table showing security and RSS solutions

S No	Security Issue	How RSS handles it?
1	Viruses, Trojans and worms on the email body.	Since the pull type of information collection is what is planned and obtained in case of RSS, it was expected that such information pertaining to the viruses, Trojans, etc., would not flow to the end user. Most of the information is authorized and therefore, cannot, ideally, have any embedded software that could possibly damage the computer they are destined to reach. XML does not carry embedded viruses unlike the document files or executables. This behaves more like the text file.
2	Viruses, Trojans and worms that come with the attachments of emails.	Viruses, Trojans and the worms spread across the network mostly as attachments to the emails. In RSS feeds also, there can be attachments. These attachments are not crosschecked by the aggregators or by the feed readers at any point of time to find out the source or the origin of these attachments. The attachments if they exist can cause the same extent of damage as they do in the case of emails. They can still distribute the viruses and the rest of their clan as much as any other email would.

Table 3: Table showing security and RSS solutions contd.

3	Sniffing on the network	RSS is still open to sniffing on the network. If the sniffer employs a key word to search for on the net and identifies an email with the password keyword, then the same is valid for the RSS feed that might go with the password. A similarly worded feed will attract as much sniffing as an email or an *http* web access might. The positive aspects of sniffing continue to exist and the negative aspects also persist. RSS is no cure against sniffing.
4	Email spoofing and the IP spoofing.	Email spoofing, if allowed a smooth run, RSS will lose its relevance. However, the communication itself might not happen through emails once the RSS is the mode of communication. Email spoofing might not really be prospering. As a matter of fact, it is on the way out. The IP spoofing will also lose its teeth with RSS since most of the data will be 'pulled' by the people who are in need of it. Therefore, the chances that some one will enact as another site does not exist unless the site itself is hijacked; in which case, it becomes a major offence/crime. RSS secures to a reasonable extent on both email spoofing and IP spoofing.
5	Splogging	Spams from the emails would have gone down drastically with the rise in RSS feeds. The user will download only that information, which is required by him and this would encourage depletion of spam guards. However, the spamming of the blogs by itself has gone up and this again leads to increased RSS feeds that are directly linked to these blogs. This would in turn increase the traffic and therefore, congestion of the networks. Spam blogging also spreads messages that are not solicited and results in waste of time and energy. RSS has little or no effect on these Splogging activities on the net.

Table 3: Table showing security and RSS solutions contd.

6	Intrusion on Private networks	One of the major scourges of the internet is the intrusion into the private space of the individual by unsolicited people. This could be controlled only if the entry is detected and stopped; and the passwords of the authorized users are appropriately protected. By using RSS, the information cannot possibly be hijacked using email spoofing or such other external emailing or pushed information gatherers. It is, however, possible for the information to flow out of the computer using any of the other methods. Using RSS cannot stop these. We can safely conclude that RSS may not be the cure for this but could control it to some extent.
7	Phishing	Phishing is another form of information or identify theft. This goes to the user in the form of an email which will take the unsuspecting victim to a website that impersonates the real one and captures the identity details for later use on the target site. This would help the hackers to gather information of the actual user and thereby hack the target system. This will not be lessened by the RSS feeds and has no relationship with the technology employed.

RSS has impact on some of the security issues particularly relating to that of the email borne security lapses. This would go down. However, it might not go down in those cases where it is not connected with the emailing activity. On others, it has serious limitations. Of course, there cannot be one cure for all the inabilities that plague the Internet.

Since this is a pull type of communication, it is easy to identify the information source, unlike in the case of emails where the source of the information on most occasions is not identifiable. If the sender of an email wants to hide, then he can do so behind a number of options that are available. But that may not be possible in the case of RSS. Therefore, it is easy to identify the source or the origin of defective data. This helps in prosecuting the defaulter if there is one. On other occasions, some of the information that comes over the RSS feeds, specifically the attachments could carry all of the above in them. They do not go through the filtering actions that the regular feeds pass through. It is important therefore, that the attachments really pass through all of the controls that the RSS feed itself passes through and the owner of the RSS feed will have to own up to the attachment lacunas as well.

While these may be viewed as major security lapses in the system and the technology, there are lapses in the individual components that make up the system. They, in turn, cause other kinds of problems and become sources of security concern to administrators.

Breach of Security in Aggregators / Readers

Readers are misled when they are made to believe that the data coming to them is valid while it is not. There was a known breach on the Yahoo's RSS aggregator vulnerability. A typical vulnerability testing was performed on the Yahoo RSS Aggregator. It was to introduce a cookie for demonstration purpose and to check whether it could get passed over to the users without really getting checked out. **[9]** It was found that this is happening even on sites like Yahoo that is very conscious of the security issues plaguing the RSS world. The experiment could pose the change on the web site as if it is originating from a regular, very common looking heading. It could have come from CNN news but it carried a malicious link, which would take the unwary visitor to the site that the interlocutor wants.

Aggregators can also be made to accept the feed in its entirety. This would include the description tag as a part of the entire record for the feed entry. The description tag can take a statement. The statement can be a text entry or an html code. The moment we say it can take html code then a whole lot of possibilities are opened up.

HTML can accept scripts, ActiveX objects and many other CSS styles as well. Now, they can do a lot of damage to the destination computer. **[10]** The statement in the description can be an HTML file that can cause all these and more. This bug in the technology front of the RSS is allowed to be passed by most aggregators. The scripts in the feed cannot only be in the description but also in the other areas. A JavaScript can be included in the main content of the feed itself. This could lead to the feed triggering off a routine; it is downloaded or possibly used to populate a web site. **[11]** The virus or a Trojan built into it could easily activate a number of other unwanted chain reactions down stream of this built up. This loophole leads to injection of the JavaScript that could install cookies in the target computer and could also steal cookies, which could behave as if it is from a different site.

A typical attack condition is illustrated below:

In our example, let us take an interested party who wants to attack a financial site. Many of the sites log on to their site, by reading their IP address. Many of them provide an always log feature that allows the users to just get into the site to have themselves logged on to it rather than typing out the log in details. If the interested party is trying to break into one such web site, then he could possibly do something like this. He could embed a malicious JavaScript code requesting for the execution of an order on the finance site.

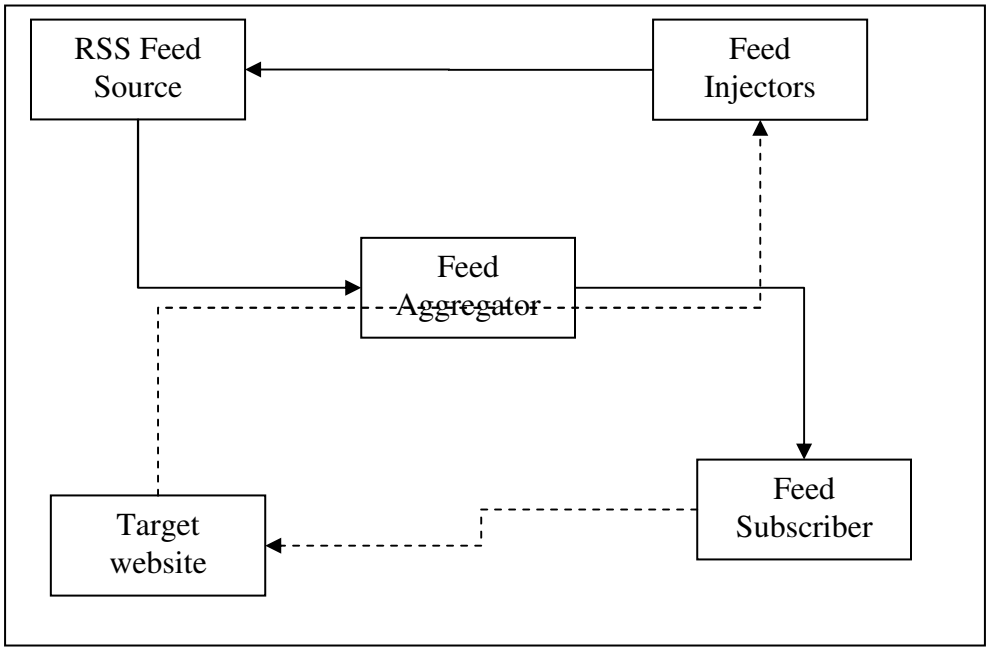


Figure 12: A typical operation tied to a feed aggregator

The script will be introduced into the feed source either as a part of the content part or in the description as a redirection to the target site with the clear indication of the order to be executed. Now, in figure 12; this insert will pass through the feed aggregator and to the feed subscriber. When the feed subscriber sees the information, the target site receives the redirection and the feed injectors get the result that they wanted.

```
<img
src="http://www.mystocktradersite.com/transaction.asp?sell=
google&buy=Microsoft&numshares=1000">
```

Figure 13: A typical html code

Embedding this html in figure 13; into the feed could result in a posting that would appear as if it originated from the feed subscriber but would have resulted in the gain of the feed injector. In order to get around this problem, the aggregators should strip the tags in the sentences that are submitted to them before resending them to their clients. There are many aggregators that are made with the needed stripers to remove these extra appendages. Only then do these aggregators become safe to operate and safe to use for the subscribers.

In many cases, the attacker could also target at the feed readers. Feed readers come in two forms; either as online or as local clients. In cases where the client is using the online readers, all the effects of the aggregators and the remote problems that are noticed and described earlier is applicable and could affect the user and the targeted system.

In the case of local clients, this effect is minimized but then the problem does exist if the client is online. And other types of issues prop up when the client is on local machines.

Typical local client side risks include cookie thefts and other types of cross-site scripting attacks. In some cases such feeds are further distributed to different users, which spread its tentacles to a number of other users too. **[12]**

Misuse of Technology

Technology of scrapping, which is normally used by third party vendors of RSS feeds, tends to misuse the results. The scrapping exercise of the screens tends to provide unauthorized copies of the data that is available on the web sites. This, in turn, is illegal and on many occasions results in copyright violations. RSS feeds and scrapping technology together spreads such copyrighted information through out the net.

Syndication is about sharing information with other people. However, there might be cases where sharing of information that are not supposed to be shared causes quite some issues. For instance, most of the blog information gets posted all over the net by using the RSS posts done from the respective blog. **[13]**

Some of these blogs are copyrighted and they might not view such actions in line with the terms of the blogs.

Scrapping as a technology was originally made for picking up changes in a site when they occur. It was for the positive purpose and could provide for the changes that were happening in the website to interested parties which is not provided for by the owner of the site. However, most of the scrappers today make use of the feed thus obtained, for advertising and their own site promotion purposes, which leaves the owner of the site loosing out on all counts.

When does a feed become malicious?

The feed becomes malicious due to a number of reasons. Some of them are listed below:

1. Malicious feed owners could be the cause of the feed becoming malicious. But this seems to be a highly improbable occurrence. Since the feed subscriptions are done only in the case of dependable origins by the client himself. Therefore, the chances of owner of the feed being malicious, is highly remote. However, there is a chance and that needs to be considered as well.

2. A site was successfully attacked and the attacker could successfully implant his feed into the sites' regular feeds. This is much safer for the attacker and will not be deducted for a long period since the effect might not be noticed and then also it has to be retraced back to the source of the feed, which is again very remote. This makes the RSS based attacks more complex and difficult to investigate.

3. Most of the malicious codes on feeds originate from common posting locations like blogs, mailing lists, bulletin board services, etc. These are the places where the administrator needs to be doubly careful on the nature of feeds that are fed to the aggregators. Of course, the possibilities of injecting such a malicious code are much more easily done in blogs than anywhere else.

4. On many occasions, the injected feeds are used as a deployment vector, which is, later on, used for cookie capturing etc.

5. There are other not so malicious damages but yet disturbing trends, which are used to cull information from the screen, using screen-scrapping techniques. Here, there is the feed owner who is at fault and third providers of feed normally do this. This results in providing information or articles that are copyrighted and in violation of such copyrights. This also causes a major loss to many of the web site owners. It is therefore, essential for all the feed providers to ensure that they have the permission of the web site administrator to make use of the needed information.

Most of the malicious code that comes with the XML feed is embedded as an html and either links to another program elsewhere on the web or brings the embedded file along with it. Therefore, if the malicious codes need to be stopped, then it is essential that these html code embedded in the XML be removed. Once this is removed most of the embedded code in the html is removed. But then, this does not do anything against the copyright violations by the person responsible for the content.

These incidents if they occur can be traced to the source of such irresponsible violation of the copyright. This possibility to trace back the source would help curtail such incidents from happening. By culling out the html feed from the XML feed, it will also help in reducing the amount of embeds that go into the feed.

Countering RSS Feed Issues

The RSS feed related issues have been controlled and curtailed to a great extent. However, the issues discussed earlier still remain. The feed by virtue of its technology, avoids most of the spam kind of problems.

Spamming the feeds with unwanted information is almost non-existent. Since any such spam is traceable back to the owner of the feed and this could result in the feed getting black listed by the users, the owners may not risk such behavior. Therefore, the chances that any body might spam the feed are countered by the technology itself.

As we have seen earlier, embedding html code within the XML code could cause such information and programs getting forward to uninterested clients. This can either be from the owner of the feed or by any other third party who might be injecting such code into the feed. In either of the case, this can be stopped by suitably updating the aggregator to filter the unwanted code from the feeds. This would make the feed free of such risks and would make the feed more secure.

Codes may be inserted at all points in the feed as we have seen elsewhere. This would also be removed by using intelligent aggregators that could different XML from html. Most of the aggregators instead of removing the html code would only remove the html

tags that are embedded in the XML. This would render the code embedded in the XML harmless.

RSS feeds are not replied to. Therefore, the harm of interjecting or sniffing out any password from or using RSS is remote. This can only happen through phishing attacks which could of course be carried by RSS feed as much as an email. In case of RSS based phishing attacks, the attacker is very well identified since this has to be the same as the owner. In case if a code was injected into the feed, then this has to be removed by the aggregator after the injection is identified. All these methods would ensure that the phishing attack is again unsuccessful.

Finally, the splogging is hardly stopped by RSS feeds. Since the happening is in line with the technology that is adopted by the user and the client has requested for certain information. The spammed blogs would generate feeds as much as any other blog would. Therefore, it is not possible to control or identify splog at this point of time and this would continue to exist.

References:

[1] CERT, 2007, CERT / CC Statistics, Carnegie Mellon University, available at: http://www.cert.org/stats/cert_stats.html

[2] Wikipedia, 2007, Attack Tree, available at: http://en.wikipedia.org/wiki/Image:Attack_tree_virus.png

[3] Tony Bradley, 2007, Introduction to Packet Sniffing, about. COM, available at: http://netsecurity.about.com/cs/hackertools/a/aa121403.htm

[4] IBM Corp, 2007, Internet Security Systems, IBM Internet Security Systems, available at: http://www.iss.net/security_center/advice/Underground/Hacking/Methods/Technical/Spoofing/default.htm

[5] Jonathan Bailey, 2006, Is RSS the problem? Plagiarism Today, available at: http://www.plagiarismtoday.com/?p=254

[6] Jon Swartz, 28 Dec 2005, 2005 Worst year for breaches of computer Security, USA Today, available at: http://www.usatoday.com/tech/news/computersecurity/2005-12-28-computer-security_x.htm

[7] Russel Kay, 19 Jan 2004, Phishing, ComputerWorld, available at: http://www.computerworld.com/securitytopics/security/story/0,10801,89096,00.html

[8] Microsoft, 2007, Recognize Phishing Scams and fraudulent e-mails, available at: http://www.microsoft.com/athome/security/email/phishing.mspx

[9] Jeremy Moeder, 1 Jan 2007, Yahoo RSS XSS Vulnerability, available at: http://addict3d.org/index.php?page=viewarticle&type=security&ID=5064

[10] Mark Pilgrim, 12 Jun 2003, How to consume RSS safely? Available at: http://diveintomark.org/archives/2003/06/12/how_to_consume_rss_safely

[11] Niall Kennedy, 4 Aug 2006, Black Hat presentations exposes RSS and Atom risks in the wild available at: http://www.niallkennedy.com/blog/archives/2006/08/black-hat-prese.html

[12] Robert Auger, 1 Jan 2007, Feed Injection in Web 2.0 available at: http://www.spidynamics.com/assets/documents/HackingFeeds.pdf

[13] Mike Rundle, 16 Jan 2006, Top Ten Sources: Stealing Your Content? Available at: http://businesslogs.com/reputation/top_ten_sources_stealing_your_content.php

RSS Security for Healthcare Communications

Health care communication is switching to more secure communications from their usual email based communication systems. Most of the information syndication in the health care is switching over to RSS feeds. **[1]** RSS feeds offer a cheaper alternative to protecting information and in sharing information with the appropriate people.

Data is shared in case of healthcare industry only with the relevant people who have subscribed to it. RSS is able to take care of maintaining feed to those people who have subscribed to the information. This makes relevant sharing of information possible and possible information duplication is avoided.

SECURITY IN HEALTHCARE COMMUNICATIONS - DATA COLLECTIONS

A questionnaire has been prepared to collect the feedback from the users of the RSS feeds to identify the usage and problems faced by users. The questionnaire is given in the appendix 1.

The questionnaire is structured in the form of a Likert scale where the users may tick off the possible alternatives of 1 to 5. This will help in easy analysis of the responses. The questions are framed into three major groups primarily aimed at providing space for easier analysis.

The first three questions in the questionnaire help in identifying the extent of usage of the RSS feeds for the user. How far he has subscribed to them and the number of times he sees the feeds that he receives and the usefulness of the feeds from them.

Out of the next set of questions, question number 5 and 8 are towards ease of use and the usability of the administrator. This would reveal to what extent the user is inclined to use RSS feeds for his own clients. This would also indicate the usefulness and the security of the feeds as well. This, he would do only if he is convinced of its usefulness and of its success in providing the needed information.

The rest of the questions are on the security aspect of RSS. Between them, they would help in understanding the respondents' inclination to create a better and more promising RSS feed that would be secure for himself and for his clients.

Importance of RSS in Healthcare communications

The use of RSS for syndication of information is catching up in Healthcare. Typical information shared includes job opportunities, information on new products and cure specifically the ones that get on to the sites of Medical companies. There are also continuous and on going up date of research information that has to be shared. This syndication is also effectively done using RSS feeds.

Most of the medical companies and their research establishments provide feeds for the doctors. The RSS feeds have to be secure without any embedded codes in them. Under a number of conditions, RSS feeds are found to be more secure than the emails or any other electronic communication formats. Example of the latest publisher is called "MedReader" This is a RSS Reader and RSS Directory specifically built for the Medical & Healthcare Professional. MedReader allows information to be directly delivered to subscriber without the worries of Spam, searching for hours, or registering/logging into a web based system. MedReader provides: **[2]**

- A built in podcast system
- Persistent search technology
- Easy to read displays for healthcare news
- A better way to manage healthcare content
- Searchable directory of Medical RSS feeds

In Healthcare communications, RSS works with the opt-in philosophy, that is if an article/site is of interest, the content could be subscribed to usually for free. MedReader can then view this information and search for more RSS content at subscribers'request. Essentially, RSS desktop readers provide a platform for people to have an opportunity to get information without registrations, passwords, or logins (except for paid subscription sites).
MedReader, a RSS News Reader and Directory built for the Medical & Healthcare Industry, is a RSS reader that not only allows Healthcare professionals to view the information, but can allow subscribers to manage, organize, and add content as could be seen fit. Additionally, MedReader uses a persistent search technology that automatically updates Medical Professionals with relevant information that will be useful.

Final Recommendations

Use of RSS feeds is on the increase. Safe and cleaner use is what is to be solicited from the users. Most of the regular users tend to make their life easier by enhancing their own appeal to the various clients by creating RSS feeds of their site. This would enable the clients to know the changes and announcements happening in the site. New products are easily identified and features are distributed within a reasonable time frame across the entire cross section of the customers. RSS feeds are therefore, here to stay.

In order to make them more secure, the feed creators or the feed owners need to ensure that there is no injunction of extraneous matter into their feed. This can be assured by providing for clear filters at their end to send out only XML feeds to the aggregators.

Secondly, the aggregators themselves should also have the capability to filter out the html code that could possibly be embedded within the XML code. This would save the feed from most of the unwanted usages. This would also save lots of trouble for the feed owner as well as to the client.

Thirdly, the feed creator should also ensure that blogs are clearly classified and spam blogs are identified and isolated. Feeds should be created away from these splogs and they should not be affecting them any further.

Fourthly, the screen scrapping software should not be used to capture information from those web sites that do not allow usage of content with out appropriate copyright permissions. Subscribing to such contents should also ensure that the aggregator has sufficient rights for such usage. This will ensure copyrighting and other rights of the owner of the content, which should be protected by every user. This would be the responsibility of the end user. These methodologies would save the clients from almost all of the issues that afflict RSS feeds as of date.

Conclusion

From our analysis and study on the RSS feeds and the technology driving it, it was found that the technology is inherently the pull type of information syndicating which would enable the user to identify and pick information only from the sources known to it rather than from all sources. This would also ensure that the information so obtained would be more dependable and trustable since it comes from the site that is known to the user and he trusts the information.

Once all these steps to enhance the security of the RSS feed is made then the work on the part of the client would greatly come down. Standard formation for RSS is being improved and this would also include the standards that aggregators should match and information sharing in the web would follow. While this would ensure swifter database collection and information dissemination across the entire gamut of users, this would also ensure that the information is given to the right person.

RSS is to ensure that the right information goes to the right person for the right kind of application. This would ensure that no or very minimal misuse of information happens over the net. This would also ensure that unwanted Trojans and viruses do not get passed around, occupying computers and bring them down. Such downtimes could get reduced more and more with people depending on RSS produces standard information packs and distribution of such information is not through emails.

Safe and secure RSS feed will always work for the advantage of the web community. It is only a matter of time before all these issues that have been raised here are addressed and the feed is made more resolute and safe for working.

References:

[1] Smith E, Eloff JHP, Apr 1999, Security in Health care Information systems - current trends, International Journal of Medical Informatics, Elsevier. Vol 54, No.1, pp 39-54.

[2] http://www.medreader.com/News/jounal

Appendix 1

Questionnaire for the Feedback on the RSS security from Users

Name		Company	
Date		Current responsibility	

S No	Description	Not at all	No	May Be	Yes	Very Much
1	Have you been using RSS feeds?	☐	☐	☐	☐	☐
2	Do you regularly see the feeds that you receive?	☐	☐	☐	☐	☐
3	Do you think RSS feeds are useful for your business?	☐	☐	☐	☐	☐
4	RSS feeds are secure. Do you accept this point?	☐	☐	☐	☐	☐
5	How time consuming is it in setting up and using?	☐	☐	☐	☐	☐
6	RSS has made communication more secure. Is this right?	☐	☐	☐	☐	☐
7	Are there loop holes in RSS that has shown up during your usage?	☐	☐	☐	☐	☐
8	Will you use RSS to pass information to your clients?	☐	☐	☐	☐	☐

What security issues do you think still persist in RSS?	
Signature	

Appendix 2:

Sample RSS feed


```
<?xml version="1.0" encoding="utf-8" ?>
- <rss xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:content="http://purl.org/rss/1.0/modules/content/" version="2.0">
- <channel>
  <title>The Health Care Blog</title>
  <link>http://www.thehealthcareblog.com/the_health_care_blog/</link>
  <description />
  <language>ar</language>
  <lastBuildDate>Tue, 06 Feb 2007 00:38:00 -0800</lastBuildDate>
  <generator>http://www.typepad.com/</generator>
- <item>
  <title>TECH/HOSPITALS: Cisco healthcare briefing</title>

    <link>http://www.thehealthcareblog.com/the_health_care_blog/2007/02/
    techhospitals_c.html</link>
  <guid
    isPermaLink="true">http://www.thehealthcareblog.com/the_health_care_bl
    og/2007/02/techhospitals_c.html</guid>
  <description>The video of the briefing on health care hosted by Cisco is up
    here. To get an idea of what it was about take a look at this agenda. Then
    go take a look at my little part of the...</description>
  <category>Technology</category>
  <dc:creator>Matthew</dc:creator>
  <pubDate>Tue, 06 Feb 2007 00:38:00 -0800</pubDate>
    </item>
- <item>
  <title>POLICY: Edwards meets Schwarzeneger & uses T word-- NFIB flips
    out</title>

    <link>http://www.thehealthcareblog.com/the_health_care_blog/2007/02/
    policy_edwards_.html</link>
  <guid
    isPermaLink="true">http://www.thehealthcareblog.com/the_health_care_bl
    og/2007/02/policy_edwards_.html</guid>
  <description>I was at a conference on Saturday when the Asst Sec of HHS in
    California and the former sec of HHS in Massachusetts managed to twist
    their tongues around how they were getting to mandates, with shared
    responsibility, provider contributions,...</description>
  <category>Policy</category>
  <dc:creator>Matthew</dc:creator>
  <pubDate>Tue, 06 Feb 2007 00:13:00 -0800</pubDate>
    </item>
- <item>
  <title>PBMs: Not responsible for anything much at all?</title>
```

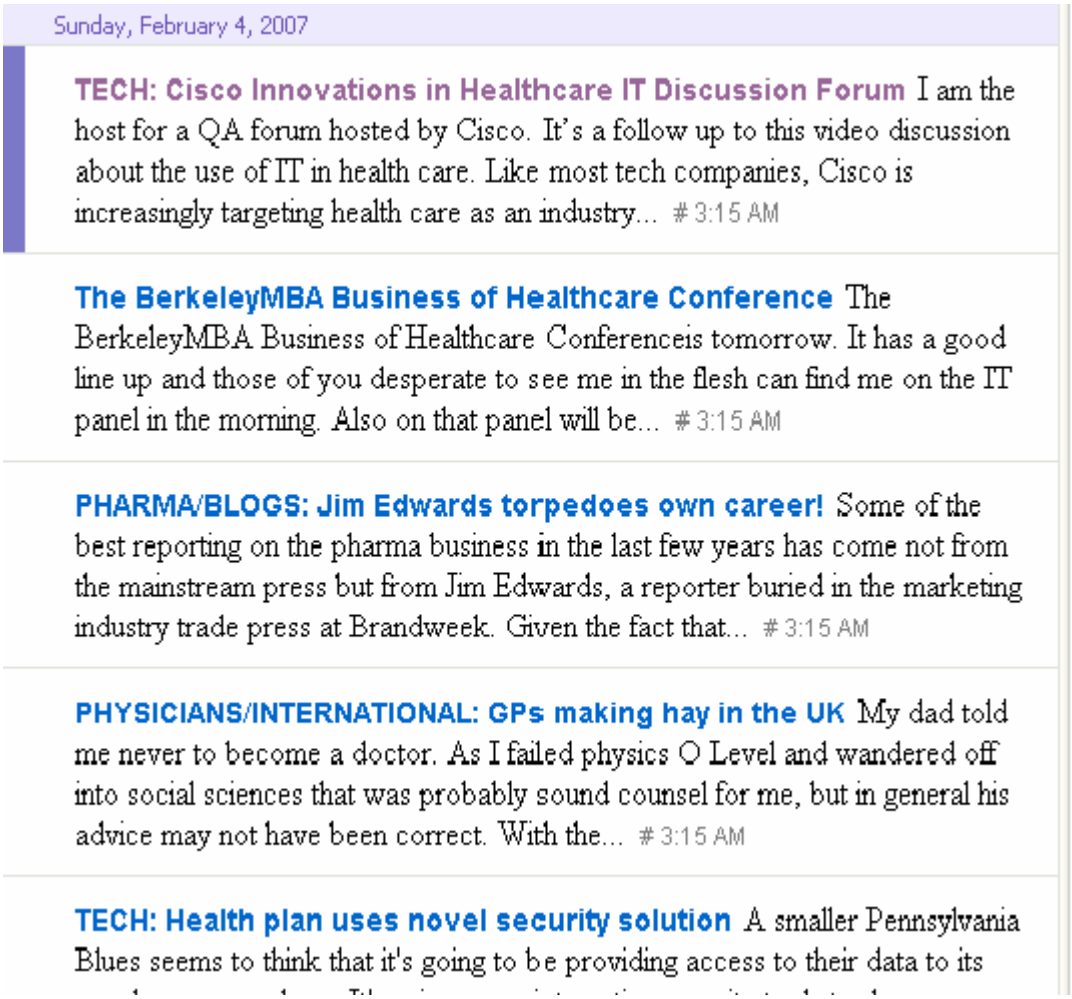
<link>http://www.thehealthcareblog.com/the_health_care_blog/2007/02/pbms_not_respon.html</link>
<guid isPermaLink="true">http://www.thehealthcareblog.com/the_health_care_blog/2007/02/pbms_not_respon.html</guid>
<description>I have purloined this and reprinted almost in full from AISHealth.com's Government News of the Week. Caremark Rx, Inc. did not breach its fiduciary duties when negotiating drug prices and managing the formulary for a multi-employer health fund because it...</description>
<category>PBMs</category>
<dc:creator>Matthew</dc:creator>
<pubDate>Tue, 06 Feb 2007 00:07:00 -0800</pubDate>
</item>
- <item>
<title>TECH/PODCAST: interview with Rahul Singal, CEO of WorldDoc</title>

<link>http://www.thehealthcareblog.com/the_health_care_blog/2007/02/techpodcast_int.html</link>
<guid isPermaLink="true">http://www.thehealthcareblog.com/the_health_care_blog/2007/02/techpodcast_int.html</guid>
<description>WorldDoc is a company that sells an interesting mix of a consumer web tools based on PHRs, care management software and transparent PBM services. Its current customers are employers and regional TPAs HMOs. I spoke with CEO Rahul Singal about...</description>
<category>Podcasts</category>
<category>Technology</category>
<dc:creator>Matthew</dc:creator>
<pubDate>Mon, 05 Feb 2007 01:03:00 -0800</pubDate>
<enclosure url="http://www.thehealthcareblog.com/the_health_care_blog/files/rahul singal.mp3" type="audio/mpeg" length="8968128" />
</item>
- <item>
<title>POLICY/INTERNATIONAL: A split in the libertarian right? (albeit a Canadian one)</title>

<link>http://www.thehealthcareblog.com/the_health_care_blog/2007/02/policyinternati.html</link>
<guid isPermaLink="true">http://www.thehealthcareblog.com/the_health_care_blog/2007/02/policyinternati.html</guid>
<description>Buried at the end of a rant about the evils of the Canadian system from our northern brethren's version of Cato/PRI—the Fraser Institute—is their solution for what to do about it all. Canada should adopt a system like Switzerland's that...</description>
<category>International</category>
<category>Policy</category>
<dc:creator>Matthew</dc:creator>
<pubDate>Mon, 05 Feb 2007 00:53:00 -0800</pubDate> </rss>

Appendix 3

Screen Shots



Page 1

Appendix 3 contd

Page 2

(The whole feed is about five pages of similar information)