WinstonSoft Weekly Report 2.9

This week:
-Worked on YMSGMatcher class.
-Worked on SVM part of the project.
-The class for NNTP protocol identification is added.
-Testing and some bugs handled.

Next week:
-Further work on YMSGMatcher class.
-Further work on SVM.
-Further work on NNTP.
-Testing and debugging.

Can HOSGOR: This week, I did some modifications to the code and worked to make it ready for the second demo . I changed SVM module so that it no longer considers bigram frequencies that are lower than a certain treshold. This allowed us to exclude some rarely transmitted (and possibly corrupted) data from SVM training dataset. I also experimented with the SVM parameters and decided to use the "linear kernel" which gave more accurate results and is somewhat faster than the default "radial basis function" kernel. In addition to the SVM module, I worked on the decoder module and fixed some issues (for example, not being able to open more than one pcap file) in it. Finally, I and İlker replaced the old toolbar icons. Next week, I'm going to team with my friends and complete the second milestone requirements.

Elvan GULEN: This week, I continued to work on the YMSGMatcher class. I analysed YMSG-WhiteBoard.doc and found out much more information than that I found in the previous week while observing some YMSG pcaps. In the summary pane the keys and values were being displayed but I organized them to show some apprehensible data like the dialog messages, status changes, file tansformation information and etc. There are some bugs about the class that I haven't handled yet but I'm planning to handle them next week and develope the log part to show more data. Besides, I will do further testting to measure the accurency of YMSG and POP3. If time allows, I will also help Can with the SVM.

Nazif Ilker ERCIN : This week, I tried to debug the program that is written up to now. I tested SMTP, POP3 and FTP matcher classes. The classes I mentioned, sometimes get lower match values for the connections related to them and sometimes get high match values for some unrelated connections. I tried to find out why these happened and fixed it in SMTP and FTP matchers. Additionally I studied on SVM and its possible usages on the project and I helped Can with the SVM module. I also studied IMAP rfc again because as a group, we decided to extend our program to identify more protocols. Next week, I'm planning to implement the IMAP matcher class and integrate it with the program. If needed, I will be dealing with SVM class. I am also planning to do some improvements on the user interface and try to implement a useful feature to filter the summary according to a key word.

Çağla ÇIĞ : This week, I spent all of my time on the implementation of NNTPMatcher class. First, I worked on the related RFC documentation and made a brief summary to help me through the implementation. Actually, with the help of this summary and the know-how about the command-response couples, the implementation was straightforward. Also, since our project is well-structured into seperate classes, the addition of a new protocol to the group of available protocols to be identified was easy. After I implemented NNTPMatcher class, the corresponding changes have also

been made in the GUI and the class is finally integrated to ACCIPP. Following that, using WireShark and telnet (telnet forums.borland.com 119) I created various pcap files consisting of NNTP commands used in different variations. Using only one of these pcaps, a simple one, ACCIPP can identify a pcap consisting of only NNTP commands with 100% accuracy. Next week, I am going to work on testing, refining and improving of this class and also cleaning bugs if found during the testing process using the pcap files I created this week. In addition to this I have found the extensions to the NNTP at http://www.ietf.org/rfc/rfc2980.txt; so next week I am also going to add these command-reply couples to my code so that the NNTP identifier more reliable.

The brief summary that assisted my through the implementation of NNTPMatcher is as follows:

When used via Internet TCP, the contact port assigned for NNTP is 119. It is a text-based protocol and commands and replies are composed of characters from the ASCII character set. Commands and command parameters are not case-sensitive. Each command line must be terminated by a CR-LF (Carriage Return-Line Feed) pair, therefore the EOL for this protocol is "\r\n". Text is sent as a series of successive lines of textual matter, each terminated with CR-LF pair. A single line containing only a period (.) is sent to indicate the end of the text. Status response lines begin with a 3 digit numeric code which is sufficient to distinguish all responses. The commands are: ARTICLE, BODY, GROUP, HEAD, HELP, IHAVE, LAST, LIST, NEWGROUPS, NEWNEWS, NEXT, POST, QUIT, SLAVE, STAT