WinstonSoft Weekly Report 2.12

This week :

- Some bugs are fixed with YMSG and the rule based classifier for YMSG is finished.
- Decoder file is analyzed.
- SVM classification is initiated.
- Some arguments for optimization.

Next week :

- Implementation for optimization.
- Working on SVM classification.
- Starting to work on Decoder.exe integration.
- Database module will be initiated.

Nazif Ilker ERCIN : This week we as a group have done a meeting with Sevgi Yasar from Siemens company. We have talked about the future progress of our project. In addition to this we have asked for a manual for the decoder supplied by Siemens. I have read the manual and tried to understand how the decoder works. Additionally I have tried to do some improvements on our program so that it works faster for the real time working issues. Next week I am planning to work on the decoder supplied by Siemens and start doing the necessary modifications in order it to be usable with our project.

Elvan GÜLEN : This week, again I worked on YMSGMatcher class. I handled the bug that I mentioned in the previous week. It is about organizing and buffering the packet info and fitting into the same structure as server sends or the client recieves. Whenever "YMSG" is passed in a dialog or in a status mesage or etc, the program couldn't catch those mesages and treated them as a new server/client post. This bug was decreasing the accuracy of the YMSG. So this week I spent some time on this bug and had to rebuild some functions. I guess from now on I'm not going to spend more time on YMSG and we aren't going to implement any other rule based classifier for any other protocols so we will focus on optimization and classification of the SVM. Also since we didn't start the implementation of the database module and the integration of the decoder.exe, we will work on these things after that. About optimization, you know generally there are too many nonmatched connections in a pcap file, and these connections decrease the performance of the program because of trying to be identified till the end of the connection data by the rule based classifiers. I and Can argued about the some solutions for this problem. Probably, we will add a feature to the GUI for faster performance and if that feature is selected than the connection data won't be stay in the rule based classifiers till the end. We believe that this will increase the performance a lot. Next week, I will work on this issue and some other optimizations. Besides I looked over the decoder document that Sevgi sent us. Maybe I will start to integrate the decoder.exe to our code. And there are some problems with the classification of SVM, if time allows I'm going to try to help Can.

Can HOŞGÖR : This week I began implementing the SVM classifier. Until this week, we provided all SVM related functionality through calling libsvm's external toolset so before going further we needed to get rid of those external dependencies. First thing I did was to to integrate libsvm code to our

project codebase, and make some adjustments in order to make it compile with Visual C++. Next, with help from some tutorials I managed to load the model files for previously trained SVMs. When the getMatchValue function is called the SVMMatcher module calls the svm_predict function with the currently collected ngrams as the feature set and receives a match value between 1 and 0. At this point the accuracy is not very good but I think it can be improved. Besides that, we had a meeting with Sevgi Yaşar this week and she told us what our top priorities should be. We are not planning to include any more protocols from now on and we are going to focus on SVM and database modules. We also discussed about performance issues and possible optimization strategies. We decided to add a condition for rule based matchers such that a rule based matcher should halt when it has a very low match value a fair amount of data has been received. Hopefully this optimization will filter out irrelevant data, not to mention it will save a lot of cpu cycles. Next week, I'm going to focus my attention on the SVM accuracy problem, and if I have time I will try to introduce some of the above mentioned optimizations.

Cagla CIG : This week, we had a meeting with Sevgi YAŞAR from Siemens. Elvan and I repeated our presentation for her and as WinstonSoft we are very glad to hear that Siemens is satisfied with our project.In the beginning of the week I started working on HTTP documentation. However, after the meeting with Ms. Yaşar, as a group we changed our minds and instead of implementing extra features to the project we decided to work on the musts of the project. Therefore, after the meeting I started working on the database module. Since the database module will be used for chart and statistical information extraction, its design is of great importance. Actually its design was finished at the end of the first semester, but it needs refinements and also new queries need to be written. In addition to this, although our project is multi-threaded it still has performance problems, so I worked on ways to improve it. Next week, I am planning to keep working on the database module and the performance problems of our project. Additionally, NNTP module still needs some minor changes so if time allows, I am planning to work on it.

WinstonSoft