Joint presentation by Niklas Lavesson and Martin Boldt
{niklas.lavesson, martin.boldt}@bth.se
Blekinge Institute of Technology
School of Computation

Spyware-hosting software has increased dramatically during the last decade. To avoid legal repercussions the vendors inform users of spyware inclusion via the End User License Agreement (EULA) during application installation. However, this information is intentionally written in a way that makes it easy to overlook. We investigate the potential to automatically discriminate between legitimate spyware- hosting software by mining EULAs. We have compiled a data set consisting of 996 EULAs out of which 10% are associated with spyware. We compare the performance of 16 learning algorithms to that of a random guesser and a state-of-the-art EULA analysis tool and show that several algorithms significantly outperform this tool. In addition, all algorithms, including the tool, performs significantly better than a random guesser (which could be used for modeling a user that is unable to comprehend the EULA). We conclude that automatic EULA classification can be used to assist users in making informed decisions about aborting or installing applications. During this presentation we will also present some new and on-going work that involves the use of machine-learning classifiers to hinder spyware.

As an alternative method for figting spyware we have also introduced the idea of a software reputation system (presented at SWITS'07). In the end of this presentation we will therefore provide an update on the current status of this work as well.
