

Intern Requirement

Preferred Discipline	<ul style="list-style-type: none"> • Computing
Prerequisites/ Skills Required:	<ul style="list-style-type: none"> • Strong interest in IT security • Programming skills would be an advantage
No. of students required	1

Project Details

Title	Automatic Malware Classification
Overview/Background	In recent years, the number of malware in the form of Internet worms, computer viruses, and trojan horses has exploded dramatically. Malware writers continue to create large number of new families and variants at an increasingly fast rate, effectively rendering manual human analysis inefficient and inadequate. Automate malware classification is fast becoming an important research in combating malware.
Aims/Objectives/Deliverables	<p>We aim to use machine learning / data mining techniques to develop a method capable of automatically classifying malware families based on information acquired from both static binary and runtime behaviour analysis.</p> <p>Deliverables</p> <ul style="list-style-type: none"> • A set of rules that can correctly identify and classify malware into relevant categories • An application that can perform static and runtime analysis of malware binaries
Scope	<ul style="list-style-type: none"> • Data collection (static binary analysis). Static analysis primarily targets the structural information of a file and exploring the characteristics of malware code e.g. to discover any use of code obfuscation techniques. A static binary-based analysis report is to be generated • Data collection (runtime behaviour analysis). For runtime behaviour analysis, malware binaries are executed and monitored in a sandbox environment. Based on state changes in the environment – in terms of API function calls – a runtime behaviour-based analysis report is to be generated • Feature extraction. Features reflecting static and behavioural patterns, such as code entry points, encryption algorithms, file structure, opening a file, locking a mutex, or setting a registry key, are extracted from the analysis reports for subsequent analysis • Learning and classification. Machine learning / data mining techniques are applied for identifying the shared pattern of each malware family. A combined classifier for all families is constructed and applied to different test data • Explaining classification. The result is analysed to yield insights into the classification model and reveal any relations between malware families and identifying potential new malware
Project Duration	<input checked="" type="checkbox"/> 6 months