



**CALIFORNIA STATE SCIENCE FAIR  
2011 PROJECT SUMMARY**

<b>Name(s)</b> <b>Ryan P. Batterman</b>	<b>Project Number</b> <b>S1403</b>
<b>Project Title</b> <b>Malware Identification by Instruction Level Code Analysis</b>	
<b>Abstract</b> <b>Objectives/Goals</b> In this project, we created models of malware and goodware (normal software) using the composing assembly instructions of executables, and then we measured the effectiveness of these models at identifying malicious code. The objective is to determine the efficacy of this new method at distinguishing the two types of software. <b>Methods/Materials</b> 53514 malware programs were obtained from the Anubis research group, and 4115 goodware were collected from a Virtual Machine created for this project. We used the Naive Bayes, Kmeans, and Support Vector Machine algorithms to create models of malware and goodware, and then we determined the effectiveness of these classifiers at differentiating malicious code from normal code. <b>Results</b> The classifiers were effective at distinguishing malware from goodware. <b>Conclusions/Discussion</b> We successfully created models of malware and goodware that differentiate the two types of software using program assembly instructions. The results indicate that this method could likely be implemented in modern antivirus solutions.	
<b>Summary Statement</b> This project investigates the effectiveness of models that use program assembly instructions to differentiate malware from normal software.	
<b>Help Received</b> My teacher Dr. Durkee read my papers; My mentor Joshua Kroll taught me about machine learning and obtained the malware set.	