|  |  |
| --- | --- |
| طرق التعدين البيانات للكشف عن البرامج الضارة | Research Title (Arabic) |
| Data mining methods for malware detection | Research Title (English) |
| البيانات والتعدين ، وكشف عن البرامج الضارة | Research Topic (Arabic) |
| Data mining, malware detection | Research Topic (English) |
| جامعة سنترال فلوريدا | Publisher (Arabic) |
| UCF | Publisher (English) |
| 2008 | Publishing Year (Arabic) |
| 2008 | Publishing Year (English) |
|  | ISBN |
| استخراج البيانات، كشف عن البرامج الضارة، التصنيف الثنائي، تحليل ساكنة، التفكيك تسلسل التعليمات، | Key Words (Arabic) |
| Data Mining, Malware Detection, Binary Classification, Static Analysis, Disassembly, Instruction Sequences | Key Words (English ) |
|  | Journal Name, or (Conference + place and date being held) |
|  | Volume No. or Issue No. and the Number of Pages in case it has been published in a scientific journal |
| هذا البحث يحقق في استخدام أساليب استخراج البيانات لالخبيثة (البرامج الضارة) كشف واقترحت إطارا بديلا للطريقة التقليدية توقيع الكشف. النهج التقليدي باستخدام التوقيعات للكشف عن البرامج الخبيثة فشل في القضية ملورس جديدة وغير معروفة، حيث التوقيعات ليست متاحة. نقدم إطارا استخراج البيانات للكشف عن البرامج الخبيثة. جمعنا وتحليلها ومعالجتها وعدة آلاف من الخبيثة برامج نظيفة لمعرفة أفضل الميزات وبناء النماذج التي يمكن أن يصنف برنامج معين في برامج ضارة أو فئة نظيفة. أبحاثنا يرتبط ارتباطا وثيقا استرجاع المعلومات وتقنيات التصنيف ويستعير عددا من الأفكار من الميدان. استخدمنا نموذج الفضاء ناقلات لتمثيل البرامج في مجموعتنا. إطارنا استخراج البيانات يتضمن فصلين منفصلة ومتميزة من التجارب. الأولى هي التجارب يشرف التعلم التي تستخدم مجموعة البيانات، التي تتألف من عدة آلاف من العينات برنامج الخبيثة ونظيفة للتدريب ، والتحقق من صحة والاختبار ، ومجموعة من المصنفات. في الدرجة الثانية من التجارب، اقترحنا استخدام التحليل المتسلسل لاختيار جمعية وميزة استخلاص التلقائي التوقيع. مع تجاربنا، كنا قادرين على تحقيق مثل ارتفاع معدل اكتشاف 98،4 ٪ وانخفاض معدل كذلك 1،9 ٪ ايجابية كاذبة عن ملورس الرواية. | Research Abstract (Arabic) |
| This research investigates the use of data mining methods for malware (malicious programs) detection  and proposed a framework as an alternative to the traditional signature detection method. The traditional approaches using signatures to detect malicious programs fails for the new and unknown  malwares case, where signatures are not available. We present a data mining framework to detect malicious programs. We collected, analyzed and processed several thousand malicious and  clean programs to find out the best features and build models that can classify a given program into a malware or a clean class. Our research is closely related to information retrieval and classification techniques and borrows a number of ideas from the field. We used a vector space model to represent the programs in our collection. Our data mining framework includes two separate and distinct classes of experiments. The first are the supervised learning experiments that used a dataset, consisting of several thousand malicious and clean program samples to train, validate and test, an array of classifiers. In the second class of experiments, we proposed using sequential association analysis for feature selection and automatic signature extraction. With our experiments, we were able to achieve as high as 98.4% detection rate and as low as 1.9% false positive rate on novel malwares. | Research Abstract (English) |