نحو نظام تلقائي لكشف وتصنيف البرمجيات الخبيثة بناء على سلوكها

الطالبة:
سجى سالم القرشي

المشرف:
د. عمر عبدالله باطرفي

المستخلص

البرمجيات الخبيثة أصبحت تشكل تهديد كبيرا على أجهزة الكمبيوتر والشبكات. ومع انتشار استخدام الانترنت والتطبيقات الشبكية أصبح حماية المصادر من كمبيوتر وشبكات ومعلومات تحدي كبير.

تقنية اكتشاف البرمجيات الخبيثة بناء على توقيعاتها هي التقنية الأكثر انتشارا. لكن البرمجيات الخبيثة حاليا أصبحت تستخدم تقنيات لتمويه أدوات الاكتشاف، فقد أثبتت الأبحاث الجديدة أن استخدام نموذج ال Hidden Markov Modelمفيد للكشف عن البرامج الضارة باستخدام الميزات التي تعكس سلوك البرمجيات الخبيثة.

في هذا البحث اقترحنا استخدام نموذج الـ HMM للتعلم بناء على بعض السلاسل ت التي تعكس سلوك البرمجيات الخبيثة، في هذا البحث استخدمنا سلاسل الـ API call و Opcode لتمثيل سلوك البرمجيات الخبيثة ، وبناء على نتائج التعلم تم تصنيف البرمجيات الخبيثة باستخدام خوارزمية للتصنيف.

فقد اقترحنا تحسين عمل الخوارزمية للتصنيف المعروف الـ k-means ، فقد تم تحسينه باستخدام خوارزمية الـ Genetic. و قارنا نتائج التصنيف قبل وبعد التحسين ، وأيضا قارنا بين السلاسل المستخدمة وأيهما أفضل للكشف عن البرمجيات الخبيثة.

وقد أثبتت التجربة أن الخوارزمية بعد التحسين نتائجها أفضل في التصنيف و أن استخدام نموذج الHMM مع خوارزمية التصنيف المقترحة Genetic K-means يمثل تقنية لاكتشاف البرمجيات الخبيثة بناء على سلوكها بنسبة اكتشاف عالية تمثل ٩٩,٨ %.

TOWARDS AN AUTOMATED BEHAVIORAL MALICIOUS CODE DETECTION AND CLASSIFICATION SYSTEM

By Saja Salem Alqurashi

Supervised By
Dr. Omar Batarfi

Malware, or malicious code, has become a major threat in computers and in the network community. With the increased use of the Internet and application-based networks, malware detection is a serious challenge. The signature-based detection technique has been widely used as the main method of detecting malware, but with obfuscation techniques, it has failed to detect modern malware. Recent research has proven that a Hidden Markov model (HMM) is useful for malware detection using features that reflect the malware behavior. The motivation in this work is to enhance the working strategy of malware detection. In this study, the related problem of malware clustering based on HMM is considered.

In meeting this goal, this study has proposed a system of testing the malware behavior based on HMM scores, which have been extracted from the learning model on application programming interface (API) call sequences and Operational code (opcode) sequence datasets as malware behavior. API call sequences that extract dynamically and opcode sequences that extract statically are used; they are compared to see which behavior is better for malware detection. Then genetic operators are used in enhancing normal *K*-means working with the HMM. The proposed genetic *K*-means is used as a classification algorithm to cluster new behaviors based on the scoring from the HMM. Next, the enhancement results are compared to the normal *K*-means classification based on the HMM, evaluating the proposed optimization technique.

Therefore, this study is considered to be an optimization enhancement and an evaluation study among normal *K*-means with the HMM in malware detection, proposing genetic *K*-means with HMM in malware detection. The results obtained from the experiments demonstrate that the objectives are successfully completed with an average of high detection rate about 87.68%.