نظام كشف التسلل المحسن للأنظمة المستضافة في المنصات السحابية

الطالب: يعقوب سيد إكرام سيد عبدالله
المشرف: أ.د. محمد اشرف اسماعيل مدكور

المستخلص

لكشف الهجمات الإلكترونية الغير معروفة مسبقاً والتي تستهدف الأنظمة الحديثة والمستضافة في المنصات السحابية، تم اقتراح عدة أنظمة مبنية على أساس المضيف للكشف عن المتسللين وذلك باستخدام مجموعة بيانات ADFA-LD التي تم تجميعها حديثاً. تقوم هذه الأنظمة المقترحة بكشف الانحياز عن السلوك الطبيعي للنظام باستخدام تقنيات تعتمد على تتبع آثار استدعاءات الوظائف الأساسية في نظام التشغيل من قبل العمليات في الذاكرة والتي تم تجميعها في مجموعة البيانات ADFA-LD. بشكل عام، يوجد في الأنظمة المقترحة قصور من عدة جوانب يتمثل في انخفاض دقة كشف المتسللين، وارتفاع نسبة الخطأ فيما تم اكتشافه، والاستهلاك العالي جداً لموارد النظام، وطول مدة التعلم على السلوك الطبيعي للنظام مع فقدان المرونة الكافية للاستجابة على التغيرات التي تطرأ على السلوك الطبيعي للنظام بشكل مستمر. وللتغلب على جميع هذه السلبيات وتحقيق أفضل مزيج من الدقة العالية، و نسبة الخطأ المنخفضة، والتعلم السريع للسلوك الطبيعي للنظام، قمنا باقتراح نظامين لكشف المتسللين مبنية على أساس المضيف. النظام الأول ينتفع من خوارزمية مستحدثة لاستخراج السلاسل القصيرة والفريدة فقط من استدعاءات الوظائف الأساسية في نظام التشغيل وذلك لتكوين اللمحة الخاصة بالسلوك الطبيعي للنظام. بعد ذلك، يتم استخدام خوارزمية مصاحبة لتصنيف سلوك العمليات واكتشاف أي انحياز عن السلوك الطبيعي. النظام الآخر يقوم باستخراج خصائص فريدة مبنية على التكرار والتردد من آثار استدعاءات الوظائف الأساسية في نظام التشغيل وذلك لتمثيل

السلوك الطبيعي للنظام. بعد ذلك، يتم استخدام تقنيات كشف الانحياز عن السلوك الطبيعي والشبه خاضعة للإشراف مثل support vector machines و k-nearest neighbors و k-furthest neighbors. قمنا بإنشاء نموذجين للمقترحين باستخدام لغة البرمجة جافا بناء على مجموعة البيانات ADFA-LD وذلك لمقارنة أداء النظامين. النتائج التجريبية أظهرت أن النظام الأول قد تفوق على النظام الثاني.

إلى حد علمنا، فإن ما توصلنا إليه من نتائج قد تفوق على جميع التقنيات المنشورة حديثاً من ناحية فترة التعلم على السلوك الطبيعي للنظام وكمية استهلاك الموارد. كما فاقت دقة الكشف في النظام المقترح من قبلنا تقريباً جميع الأنظمة المقترحة مؤخراً بالمقارنة وكانت الدقة شبه مساوية لأفضل ما تم نشره. وبشكل خاص، فإن نظام كشف التسلل من خلال كشف الانحياز عن السلوك الطبيعي للنظام، والمبني على خوارزمية استخراج السلاسل القصيرة والفريدة فقط من استدعاءات الوظائف الأساسية لنظام التشغيل قد جمع بين مزيج من عدة مزايا فضلى كارتفاع دقة كشف المتسللين وانخفاض فترة التعلم. لقد حقق النموذج الذي تم تطويره نسبة من الدقة العالية تساوي ٩٠,٤٨% ونسبة خطأ تساوي ٢٢,٥% مع فترة تعلم على السلوك الطبيعي تساوي تقريباً ٣٠ ثانية فقط. إن هذا المزيج من المزايا يمكّن من اكتشاف معظم الهجمات الإلكترونية الغير معروفة مسبقاً ويجعل النظام ذو مرونة ليتواءم ويتماشى مع أي تعديلات في البيئة نظراً لأنه قابل لأن يتعلم السلوك الطبيعي الجديد بسرعة، وبشكل تكاملي من دون الحاجة إلى بناء كامل خوارزمية التصنيف من الصفر.

Enhanced Host-based Intrusion Detection System for Cloud Platform

Student: Yaqoob Sayedikram Sayedabdullah
Supervisor: Prof. Mohamed Ashraf Ismail Madkour

ABSTRACT

To detect zero-day attacks in modern cloud platforms, several host-based intrusion detection systems are proposed using the newly compiled ADFA-LD dataset. These techniques use the system call traces of the dataset to detect anomalies. The common limitations found in such systems include one or more of the following: low detection rate, high false alarm rate, and long learning time that leads to inflexible response to eventual changes in the normal profile. To overcome these limitations and achieve best combination of high detection rate, low false alarm rate, and small learning time, we propose two host-based intrusion detection systems. The first system utilizes a novel algorithm to extract only distinct short sequences of system calls per normal trace to create a normal profile. Then, a companion classification algorithm is used to detect anomalies. The second one employs frequency-based feature extraction from traces of system calls and uses semi-supervised anomaly detection techniques such as support vector machines, k-nearest neighbors and k-furthest neighbors. We developed two prototypes using Java language for both systems and compared their performance using the ADFA-LD dataset. The experimental results showed that the first system outperformed the second.

To the best of our knowledge, the obtained results of the proposed first system are superior to all up-to-date published systems in terms of computational cost and learning time. The obtained detection rate is also much higher than almost

all compared systems and is very close to the highest result. In particular, the proposed short-sequence-based intrusion detection system provides the best combination of high detection rate and very small learning time. The developed prototype achieved 90.48% detection rate, 22.5% false alarm rate, and a learning time of about 30 seconds. This provides high capability to detect zero-day attacks and also makes it flexible to cope with any environmental changes since it can learn quickly and incrementally without the need to rebuild the whole classifier from scratch.