البرامج الضارة (اختصار للبرامج الضارة )، بما في ذلك الفيروسات والأبواب الخلفية وبرامج التجسس وأحصنة طروادة والديدان وشبكات الروبوت، في شكل تعليمات برمجية قابلة للتنفيذ، يمكن استخدام عينات البرامج الضارة لتعطيل تشغيل الكمبيوتر، يعتمد النهج الرئيسي للحماية من البرامج الضارة على التوقيع والذي يتم اعتماده على نطاق واسع من قبل معظم بائعي برامج مكافحة البرامج الضارة [8]، التوقيع هو جزء معين من التعليمات البرمجية يتم التعبير عنه في شكل بايت أو تسلسل تعليمات يتم الحصول عليه من خلال التحليل اليدوي بواسطة خبراء أمن الكمبيوتر، وهو فريد لكل برنامج ضار معروف يسمح بالتعرف على الملفات غير المعروفة بشكل صحيح مع معدل اكتشاف أعلى [10 ]. يتم نشر المراقبة وتحليل التعليمات الثنائية لاستخراج التوقيعات من عينات البرامج الضارة غير المعروفة يدويًا، مدفوعين بالفوائد الاقتصادية، بما في ذلك التشفير وتعدد الأشكال والتحول لجعل عينات البرامج الضارة تبدو وكأنها تتصرف مثل الملفات الحميدة وتكون محصنة ضد الاكتشاف القائم على التوقيع، لقد حفزت هذه المشكلة صناعة مكافحة البرامج الضارة على تكريس نفسها لإعادة تصميم أنظمة الأمان الخاصة بها لاكتشاف البرامج الضارة التي تم إصدارها حديثًا. وفي مواجهة الكثير من عينات الملفات غير المعروفة حديثًا كل يوم، تم إجراء العديد من الجهود مع تقنيات استخراج البيانات لتحليل وتصنيف عينات الملفات، طبقت هذه التقنيات خوارزميات استخراج البيانات (مثل التصنيف أو التجميع) لاكتشاف البرامج الضارة استنادًا إلى ميزات المحتوى، بالإضافة إلى محتوى الملف، يتم أيضًا الاستفادة من العلاقات بين عينات الملفات (مثل العلاقات من ملف إلى جهاز، والعلاقات من ملف إلى ملف، والعلاقات من ملف إلى أرشيف) لاكتشاف البرامج الضارة في السنوات الأخيرة [5]، تم تطبيق خوارزمية نشر التسمية على الرسم البياني لعلاقة الملف الذي تم إنشاؤه لاكتشاف عينات الملفات الضارة. فمن المستحيل وغير المقبول أن يقوم خبراء الأمن بتحليل جميع هذه الملفات وتسميتها بسبب التكلفة الباهظة. دورًا مهمًا في تقليل تكلفة وضع العلامات وتوفير التحسين الأكثر أهمية في نموذج للكشف عن عينات الملفات الضارة FindMal التعلم. وبالاستناد إلى الرسم البياني للعلاقة بين الملف والملف، نقدم إطار عمل والتعلم النشط. بدلاً من استخدام معلومات محتوى عينات الملفات، فإننا Label Propagation التي يتم فيها استخدام طريقة نتحقق من كيفية استخدام علاقات الملفات لاكتشاف عينات البرامج الضارة وتطبيق طريقة نشر الملصقات لتصنيف عينات الملفات بناءً على الرسوم البيانية لعلاقات الملفات التي تم إنشاؤها. تم اقتراح ثلاث ميزات قوية تعتمد على الرسم البياني لاختيار عينات ملفات تمثيلية للاستعلام عن التسميات من الخبراء البشريين. استنادًا إلى الميزات المقترحة والرسم البياني لعلاقة الملف، هذا هو العمل الأول لاقتراح إطار عمل لاكتشاف البرامج الضارة استنادًا إلى الشبكة الاجتماعية لعينات الملفات باستخدام خوارزمية التعلم شبه الخاضعة للإشراف مع أساليب التعلم النشط. يتم استخدام مجموعة بيانات علاقة الملفات الحقيقية والواسعة النطاق من شركة صناعة مكافحة البرامج الضارة في التجارب. 487 ملفًا غير معروف) على 3793 عميلًا.