Streszczenie

Złośliwe oprogramowanie jest poważnym zagrożeniem współczesnego Internetu. Przestępcy używają go do wysyłania niechcianych wiadomości, wymuszania okupu przez zaszyfrowanie plików czy wykradania danych logowania do banku. Do komunikacji wykorzystywane są w nim popularne protokoły sieciowe, w tym często protokół HyperText Transfer Protocol (HTTP). Celem niniejszej rozprawy doktorskiej jest wykazanie, że żądania tego protokołu wygenerowane przez różne rodziny złośliwego oprogramowania mogą być użyte do ich identyfikacji i klasyfikacji. Do przeprowadzenia oceny eksperymentalnej stworzono zbiory danych ruchu sieciowego obejmujące 121 rodzin złośliwego oprogramowania oraz zestaw popularnych aplikacji niezłośliwych. Przeprowadzone badania podzielono na trzy części. W części pierwszej dokonano identyfikacji charakterystycznych cech żądań HTTP umożliwiających odróżnienie złośliwego oprogramowania od aplikacji niezłośliwych. Cechy te stały się bazą do drugiej części analizy, której efektem było stworzenie narzędzia *Hfinger* umożliwiającego tworzenie unikalnych reprezentacji żądań HTTP. Reprezentacje te można wykorzystać do identyfikacji złośliwego oprogramowania przez rozróżnienie jego rodzin, a także jego konkretnych działań, np. ataków lub pobierania rozkazów. W części trzeciej skupiono się natomiast na problemie klasyfikacji złośliwego oprogramowania przy użyciu algorytmów uczenia maszynowego, tzn. przypisania nazw konkretnych rodzin do analizowanego ruchu sieciowego. Problem ten został rozszerzony o rozpoznanie obecności klas, które były nieznane w trakcie treningu klasyfikatora, czyli tzw. rozpoznawanie otwartozbiorowe (Open Set Recognition). Wykorzystano przy tym dwa sposoby reprezentacji żądań HTTP: bazujący na narzędziu *Hfinger* oraz na analizie n-gramowej. Według wiedzy autora niniejszej rozprawy jest to pierwsza praca wykorzystująca rozpoznawanie otwartozbiorowe do klasyfikacji ruchu protokołu HTTP złośliwego oprogramowania.

Słowa kluczowe: złośliwe oprogramowanie, analiza ruchu sieciowego, protokół HTTP, klasyfikacja, rozpoznawanie otwartozbiorowe.