1. Opis merytoryczny

Celem proponowanego rozwiązania jest opracowanie (i dopracowanie) metod analizy zachowań użytkowników serwisów internetowych. Bazując na opracowanych algorytmach możliwe będzie zbudowanie zarówno lepszych zabezpieczeń serwisów internetowych jak i potencjalnie poprawa wydajności aplikacji – poprzez predykcję i generowanie z wyprzedzeniem części stron serwisów.

Realizacja wskazanego celu nastąpi przez opracowanie zunifikowanych metod reprezentacji zapytań HTTP, reprezentacji stanu, oraz opis przejść pomiędzy stanami. Na tej podstawie stworzone zostaną algorytmy budowy profilów użytkowników – ich grupowania i tworzenia reprezentacyjnych wzorców. Algorytmy te oparte będą na metodach analizy grafów, jako że przewiduje się iż reprezentacja zachowań użytkowników będzie przedstawiona w tej właśnie formie.

Proponowana metoda jest nowym podejściem, opierającym się na analizie zapytań użytkowników w kontekście ich całej sesji (kolejnych działań w interakcji z serwisem internetowym). Dotychczasowe metody, wykorzystywane zresztą w wielu narzędziach i aplikacjach (jak np. log watcher, octopussy itp.), opierają się z reguły na analizie pojedynczych zapytań, w oderwaniu od kontekstu, najczęściej przy wykorzystaniu statycznego zbioru reguł. Takie tradycyjne podejście skutkuje z jednej strony mniejszą skutecznością detekcji i/lub dużą ilością fałszywych alarmów, z drugiej zaś wymaga od administratora ciągłej 'ręcznej' aktualizacji w przypadku modyfikacji struktury lub funkcjonalności serwisu internetowego. W przypadku serwisów o dużym nasileniu ruchu ilość fałszywych alarmów i konieczność ich ręcznej analizy praktycznie uniemożliwia w ogóle stosowanie takiego typu ochrony.

W ramach proponowanych prac przeprowadzona zostanie analiza logów serwerów internetowych, oraz zbudowane zostaną algorytmy ich automatycznej analizy, ekstrakcji sesji użytkowników oraz ich odwzorowanie na strukturę grafu o ważonych krawędziach – odpowiadających częstotliwości przejść pomiędzy stanami. Przebadane zostaną też różne formy reprezentacji i utrzymywania wiedzy w strukturze grafowej – pod kątem możliwych zastosowań do analizy zachowań użytkowników. Oczekiwanym efektem jest segmentacja i wydzielenie typowych wzorców zachowań użytkowników. Na tej podstawie przygotowany algorytm będzie w stanie ocenić kolejne zapytanie przychodzące od użytkownika pod kątem jego zgodności z typowymi wzorcami – pozwalając zidentyfikować zapytania nietypowe, będące potencjalnym zagrożeniem (np. atakiem na serwis internetowy).

Ostatecznym rezultatem badań będzie przygotowanie i przetestowanie prototypu służącego do ochrony serwisu internetowego, potrafiącego budować na bieżąco wzorce zachowań jego użytkowników i raportować nietypowe zachowania.

2. Charakterystyka i typ potencjalnych nabywców

Potencjalnymi nabywcami końcowego produktu są wszelkie przedsiębiorstwa i organizacje utrzymujące serwisy internetowe zapewniające interakcje z użytkownikiem, choć przypuszczalnie największą wartość będzie stanowiło ono dla dużych podmiotów (ze względu na potencjalnie największe koszty możliwych nadużyć).

3. Opis istniejących materiałów promocyjnych

Na razie brak.

4. Potencjalni rozmówcy (autorytety w dziedzinie)

Wydaje się że dobrym pomysłem byłoby zaangażowanie osób z branży bezpieczeństwa serwisów internetowych (np. OWASP Poland Chapter), bądź też administratorów (bezpieczeństwa) dużych serwisów internetowych – np. portali.

5. Opis silnych i słabych stron projektu

Główną siłą proponowanego rozwiązania jest dostarczenie właścicielom/administratorom narzędzia do automatycznej analizy i raportowania działań użytkowników ich serwisu internetowego – w szczególności zachowań anomalnych. Jest to dodatkowe narzędzie podwyższające bezpieczeństwo takiego serwisu. Niestety, jak większość narzędzi z tego zakresu nie daje ono gwarancji całkowitej ochrony – pozwala jedynie zabezpieczyć system przed pewną klasą ataków – i jako takie może być elementem całego systemu bezpieczeństwa.

6. Wskazanie czynników ryzyka

Choć wydaje się że na ta chwilę nie ma istotnych zagrożeń dla możliwości stworzenia założonego systemu - jako że część algorytmów jest już wstępnie rozpoznana, to jak w każdym projekcie istnieje ryzyko że ostatecznie uzyskane rozwiązanie nie będzie zapewniało oczekiwanych korzyści. Jest to naturalne ryzyko związane z praca badawczą.

Drugim czynnikiem ryzyka (choć raczej szansą) jest kwestia możliwości pozyskania do współpracy administratorów bezpieczeństwa dużych serwisów informatycznych. Wydaje się że wykorzystanie ich wiedzy praktycznej, a przede wszystkim możliwość testowania opracowanych rozwiązań w warunkach produkcyjnych byłoby istotną wartością dla całego projektu. Niestety, jako że poruszane kwestie dotyczą bezpieczeństwa i danych wrażliwych (np. identyfikatory sesji), konieczne jest wypracowanie odpowiedniego zaufania pomiędzy stronami, co może być trudne i/lub wymagać dłuższego czasu.