Wstęp

Wstęp

Spis treści

1. Wprowadzenie do fingerprintingu **6**

1.1. Podstawowe pojęcia 6

1.1.1. Nomenklatura używana w tej pracy 6

1.1.2. Definicje 7

1.1.3. Właściwości fingerprintu 8

1.2. Fingerprinting a Internet 8

1.2.1. Podstawy funkcjonowania Internetu 9

1.2.2. Założenia funkcjonowania Internetu 9

1.2.3. Realizacja założeń funkcjonowania Internetu 9

Rozdział 1.

Wprowadzenie do fingerprintingu

1.1. Podstawowe pojęcia

1.1.1. Nomenklatura używana w tej pracy

Pisząc o odcisku palca użyto (także w tytule pracy) ogólnie przyjętego skrótu myślowego, oznaczającego odbitkę linii papilarnych, czyli formę językową uznawaną za poprawną przez specjalistów od daktyloskopii.

Użycie formy językowej „odcisk palca” w terminie „cyfrowy odcisk palca” ma wiele sensu. Jeszcze bez zdefiniowania tego specjalistycznego terminu, możemy domyślić się, co oznacza. Oczywiście wynika to z faktu, że cyfrowy odcisk palca i analogowy odcisk palca są ze sobą w pewien sposób powiązane (koncepcja cyfrowego odcisku palca czerpie z wartości wynikających ze stosowania odbitek ludzkich linii papilarnych w dziedzinie kryminalistyki).

Angielskie słowo „fingerprint” tłumaczy się jako odcisk palca, jednakże w zagranicznych publikacjach dotyczących cyfrowego odcisku palca rzadko występuje termin „digital fingerprint”. Kontekst użycia jest na tyle wyraźny, że użycie samego „fingerprint” jest wystarczające.

Zachodnie nazewnictwo ma tę przewagę, że jest zdecydowanie bardziej kompaktowe. Także w przypadku słowotwórczego zabiegu *fingerprinting*, oznaczającego czynność; szukając polskiego odpowienika musielibyśmy sięgnąć po „cyfrowe znakowanie”. Z uwagi na tę kompaktowość i łatwość użycia, w pracy preferowane będzie użycie oryginalnej nomenklatury.

1.1.2. Definicje

W kolejnych punktach zawarto najważniejsze definicje i powiązane pojęcia, które będą używane w przeciągu całej pracy.

Fingerprint

Wektor cech pozwalający zidentyfikować dowolny zbiór danych.

Aby fingerprint pełnił praktyczną funkcję identyfikacyjną, tak jak odbitka ludzkich linii papilarnych pełni praktyczną funkcję identyfikacyjną, często stosuje się algorytm, który kojarzy wektor cech z określonej długości (zwykle krótkim) ciągiem bajtów (identyfikatorem). Takim algorytmem może być na przykład wysokiej wydajności funkcja skrótu (niekoniecznie zdatna do zastosowań kryptograficznych—na przykład MurmurHash). W niektórych źródłach można także spotkać się z taką definicją, że fingerprint to już sam wynik wyżej wspomnianego algorytmu. Taka definicja nie zmienia istoty fingerprintu, ale jest mniej przydatna w kontekście fingerprintingu urządzeń podłączonych do Internetu i przeglądarek internetowych, czego dotyczy niniejsza praca.

Fingerprint urządzenia podłączonego do Internetu

Wektor cech pozwalający zidentyfikować urządzenie podłączone do Internetu.

Instalacja przeglądarki internetowej

Instalacja na konkretnym urządzeniu. W przypadku zmiany ustawień, konfiguracji i liczby pluginów oraz aktualizacji przeglądarki, instalacja przeglądarki pozostaje ciągle tą samą instalacją.

Fingerprint przeglądarki internetowej

Wektor cech pozwalający zidentyfikować instalację przeglądarki internetowej.

1.1.3. Właściwości fingerprintu

Ludzkie linie papilarne są na ogół niepowtarzalne, niezmienne i nieusuwalne. Z wartości wynikających ze stosowania ich w swojej dziedzinie badawczej czerpie (także etymologicznie) koncepcja fingerprintu i dlatego też fingerprint z dobrze dobranymi cechami będzie odzwierciedlać podobne właściwości.

W przypadku fingerprintingu urządzeń podłączonych do Internetu i przeglądarek internetowych najważniejszymi ich właściwościami są unikalność / różnorodność (niepowtarzalność) oraz stabilność (niezmienność), przy czym zwiększenie unikalności lub stabilności ma najczęściej negatywny wpływ na drugi parametr.

Jedną ze stosowanych[1] metod pomiaru unikalności fingerprintu urządzeń i przeglądarek jest entropia Shannona.

Entropia Shannona

Wartość entropii można rozumieć jako liczbę pytań binarnych potrzebnych do sklasyfikowania losowo wybranego elementu z danego zbioru. Entropia Shannona zbioru D z etykietami $\{l_0, l_1, \ldots, l_{n-1}\}$ wyraża się wzorem

$$H(D) = -\sum_{i=0}^{n-1} p(l_i) \log_2 p(l_i)$$

gdzie $p(l_i)$ to wyrażona ułamkiem częstość $x \in D$ mającego etykietę l_i. W przypadku w którym każda etykieta występuje tak samo często entropia ma wartość maksymalną równą $\log_2 (n-1)$.

Przykład: jeśli zbiór fingerprintów przeglądarek internetowych ma 32 bity entropii, to w przypadku losowego wyboru jednego z nich oczekujemy, że w najlepszym przypadku tylko 1 na 4294967295 przeglądarek będzie miała taki sam fingerprint.

1.2. Fingerprinting a Internet

Aby lepiej zrozumieć istotę fingerprintu i motywację stojącą za stosowaniem fingerprintingu w różnych obszarach techniki, kolejne punkty posłużą jako referencja (także historyczna).

[1] Metrykę tę stosowało na przykład badanie „How Unique Is Your Web Browser?" Electronic Frontier Foundation; w momencie pisania pracy jedno z największych badań tego typu.

1.2.1. Podstawy funkcjonowania Internetu

1.2.2. Założenia funkcjonowania Internetu

1.2.3. Realizacja założeń funkcjonowania Internetu

Spis rysunków

Spis tablic