2 Reprezentacja danych

2.1 Common Vulnerabilities and Exposures

Common Vulnerabilities and Exposures jest systemem zapewniającym publiczne, darmowe informacje na temat ekspozycji i podatności systemów informatycznych. Celem stworzenia CVE było utworzenie standardu, który ułatwiłby dzielenie się informacjami pomiędzy różnymi narzędziami i organizacjami zajmującymi się tematyką cyberbezpieczeństwa. Dzięki CVE możliwa stała się komunikacja pomiędzy tymi narzędziami, co pozwoliło na bardziej efektywną walkę z atakami na systemy informatyczne.

W CVE podatność zdefiniowana jest jako słabość w logice obliczeniowej, znaleziona w oprogramowaniu lub komponentach sprzętowych, które, jeśli zostałyby odkryte, niekorzystnie wpłynęły by na poufność, integralność lub dostępność systemu. Naprawa podatności zwykle wymaga zmian w kodzie programu, czasem również zmian w specyfikacji. Ekspozycją (TODO przemyśleć tłumaczenie "exposure") jest błędna konfiguracja systemu lub błąd w oprogramowaniu, który daje dostęp do informacji i zdolności systemu, użytych później przez hakera jako wytrych[2].

Lista Common Vulnerabilities and Exposures składa się CVE Entry, czyli wpisów CVE, które identyfikują poszczególne podatności.
Każdy wpis na liście CVE składa się z:

- CVE ID, który jest unikatowym identyfikatorem wpisu. Zapisany jest obecnie w formacie CVE-YYY-NNNN, gdzie YYYY jest liczbą oznaczającą rok publikacji wpisu, a NNNN jest numerem identyfikacyjnym w danym roku. pole NNNN składa się co najmniej z czterech cyfr, jednak od 2014 roku może ich być więcej niż cztery.

- Krótki opis podatności lub ekspozycji systemu, który może zwierać takie informacje, jak nazwa producenta podatnego oprogramowania.

- Odnośniki do innych, powiązanych wpisów CVE oraz wszelkie inne odnośniki, takie jak raporty i porady podane przez producenta.

Wpisy tworzone są przez zespół Mitre zajmujący się CVE, producentów oprogramowania, tak zwanych CVE Numbering Authority (CNA), których aplikacje zawierają podatności. CNA to organizacja zajmująca się dystrybuowaniem CVE ID. Ma ona ograniczone możliwości przypisywania CVE ID, dokładnie określone i udokumentowane, zwykle obejmujące programy lub sprzęt pochodzące maksymalnie od kilku producentów. Ograniczenia te stworzone są po to, by kompetencje zbyt wielu osób nie nachodziły na siebie, co mogłoby powodować chaos[3].

CVE ID jest polem wpisu CVE. Jest to unikatowy numer identyfikacyjny pozwalający jednoznacznie odróżnić od siebie wpisy na liście CVE. Składa się on z trzech podidentyfikatorów: identyfikatora listy, identyfikatora roku oraz identyfikatora wpisu w danym roku. Identyfikator listy może przyjmować dwie wartości:

- CVE - najczęściej używany, oznacza, że wpis widnieje na liście CVE
- CAN - pochodzący od angielskiego słowa "candidate", czyli kandydat, oznacza, że w momencie publikacji listy eksperci z CVE jeszcze nie zatwierdzili tego wpisu, jako zweryfikowanej podatności. Od pewnego momentu niemożliwym stało się, by zarząd CVE zajmował się każdym wpisem z osobna, gdyż było ich zbyt dużo, w związku z czym zaprzestano stosować identyfikatora CAN.

Kolejny podidentyfikator składa się z czterech cyfr i oznacza rok, w którym wpis został opublikowany. Możliwa jest sytuacja, że podatność została wykryta wcześniej, ale przy przypisywaniu CVE ID liczy się rok, w którym to ID zostało przypisane. Ostatni człon CVE ID zbudowany jest z co najmniej czterech cyfr oznaczających identyfikator unikatowy w skali roku. Od 2014 roku, w związku z rosnącą ilością wpisów na liście CVE, przyjmuje się, że ten człon może posiadać więcej niż cztery cyfry.

Opis wpisu CVE stworzony jest zwykle przez zespół CVE, organizację CNA lub osoby indywidualne zgłaszające odkrytą podatność. Opisy powinny zapewniać informacje na temat produktu w którym została wykryta podatność, identyfikatora wersji tego produktu oraz jego dostawcy. Powinien też zostać wyszczególniony typ podatności, jej wpływ na systemy informatyczne, a także informacje jak głębokiego dostępu do sytemu potrzebuje haker, by móc skorzystać z danej podatności.

Możliwe jest też wskazanie części kodu lub konkretnego komponentu odpowiadającego za podatność. Niestety, autor wpisu CVE nie zawsze jest w posiadaniu wszystkich informacji. Wynika to z faktu, że nie wszystkie z powyższych informacji są dostępne publicznie, co znacząco utrudnia wykonanie pełnego, oddającego całość problemu opisu. Z uwagi na znaczenie pola opisu z punktu widzenia niniejszej pracy magisterskiej, zostanie ono dodatkowo przeanalizowane w osobnej sekcji.

Odniesienia CVE jest polem, w którym wpisane są wszystkie istotne odnośniki do danej podatności. Mogą to być odnośniki do informacji o podobnych podatnościach, odnośniki do informacji o podatnościach powiązanych w jakiś sposób z danym wpisem CVE, jak również informacje o kontakcie do producenta danego oprogramowania lub sprzętu. Odnośniki te powinny dawać możliwość znalezienia rozwiązania poprawiającego daną podatność.

Data stworzenia wpisu, jest jak nazwa wskazuje, datą oznaczającą dzień, w którym powstał wpis. Nie oznacza ona jednak, że to w tym dniu został przypisany CVE ID. W przypadku wpisów tworzonych bezpośrednio zespół Mitre CVE, data ta pokrywa się z datą przypisania CVE ID. Jednak w przypadku wpisów tworzonych przez CNA dopuszczalna jest sytuacja, w której CVE ID jest zajęte przez daną organizację z góry, przez co nie zostaje ono użyte przez dłuższy czas. Dopiero po wykryciu konkretnej podatności w oprogramowaniu lub sprzęcie danej organizacji, wpisana zostaje data stworzenia wpisu i zaczyna on oficjalnie widnieć na liście CVE.

W wyjątkowych sytuacjach wpisy na liście CVE mogą zawierać dodatkową informację o stanie, w jakim aktualnie się znajdują. Są to stany, w których wpis nie jest jeszcze oficjalnie opublikowany na liście. Mogą to być:

- "RESERVED" - stan, w którym zajęty został identyfikator CVE ID, jednak szczegóły dotyczące danej podatności nie są jeszcze znane. Wpis może zostać opublikowany w każdej chwili, po uzupełnieniu brakujących informacji.
- "DISPUTED" - stan, w którym dana podatność jest jeszcze dyskutowana przez różne zainteresowane podmioty. Może to być na przykład sytuacja, w której zastrzeżenia do wpisu ma zespół CVE. W przypadku korzystania z wpisu oznaczonego jako DISPUTED należy, w poszukiwaniu najświeższych informacji, dodatkowo sprawdzić pole odniesień.

- "REJECT" - wpis odrzucony przez zespół CVE. Powodem odrzucenia wpisu może być na przykład powtórzenie istniejącego wcześniej wpisu.

2.2 Dane publikowane przez National Institute of Standards and Technology

National Institute of Standards and Technology (NIST) jest amerykańską agencją federalną, która zajmuje się szeroko pojętą metrologią. Wśród obszarów zainteresowań NIST znalazły się również podatności w systemach informatycznych[?]. National Vulnerability Database (NVD) jest wynikiem prac prowadzonych przez Instytut, które miały na celu usystematyzowanie i jak najlepsze opisanie podatności oraz ich wpływu na systemy informatyczne. NVD jest rządowym repozytorium, które za pomocą protokołu Secutiry Content Automation Protocol (SCAP) pozwala na ustandaryzowane automatyzowanie zarządzania podatnościami. NVD zawiera listy kontrolne bezpieczeństwa, wady oprogramowania związane z bezpieczeństwem, błędy konfiguracyjne, nazwy produktów oraz ocenę ich wpływu na system. Dane w NVD aktualizowane są na bieżąco, zaraz po przeanalizowaniu ich przez zespół ekspercki[4].

W protokole SCAP wykorzystuje się system Common Vulnerabilities and Exposures (CVE), który jest zarządzany i rozwijany przez firmę Mitre Corporation. Bazy danych NVD zawierają indeksy CVE rozszerzone dodatkowo o ocenę podatności według metryk Common Vulnerability Scoring System (CVSS). Pozwala to na podjęcie automatycznej decyzji o akcjach związanych z pojawieniem się podatności w systemie informatycznym. Baza danych NVD jest w pełni zsynchronizowana z listami CVE, co oznacza, że wszelkie poprawki i zmiany w listach CVE widoczne są natychmiast w bazie NVD. CVE znajdujące się w bazie NVD mogą znajdować się w jednym z siedmiu stanów:
Received - CVE został niedawno opublikowany do słownika CVE i został odebrany przez NVD
Awaiting Analysis - CVE został przeznaczony do analizy eksperckiej ze strony NVD, która zwykle zajmuje do 24 godzin
Undergoing Analysis - CVE jest obecnie analizowane przez ekspertów NVD
Analyzed - CVE zostało całkowicie przeanalizowane oraz wszystkie dodatkowe dane zostały do niego przypisane. Każda z analiz może być opisana

jednym z trzech podtypów:

Initial - używany by pokazać, że została przeprowadzona dopiero pierwsza analiza danego CVE

Modified - używana by pokazać, że została przeprowadzona analiza związana z modyfikacją danych w CVE

Reanalysis - używany by pokazać, że ponowna analiza została przeprowadzona z powodu innego niż modyfikacja CVE

Modified - CVE było poprawione przez źródło, które je opublikowało, co oznacza, że wyniki analizy NVD mogą być już nieaktualne

Deferred - jeśli CVE znajduje się w tym planie oznacza to, że nie jest planowana jego analiza

Rejected - CVE w tym stanie znajduje się w bazie danych NVD, nie wyświetla się jednak w wynikach wyszukiwania

Dodatkową kategorią indeksu CVE jest kategoria "Reserved". CVE występujące w słowniku, które są oznaczone takim polem, nie zostają włączone do bazy danych NVD[4].

W ramach rozwijania możliwości list CVE, baza danych NVD wprowadza nowe, zaawansowane metody wyszukiwania podatności. W National Vulnerability Database możliwe jest filtrowanie wpisów na podstawie systemu operacyjnego, na którym występuje dana podatność, czy kategoria podatności według opisu Common Weakness Enumeration. Co więcej, możliwe jest wyszukanie wszystkich podatności związanych z danym producentem lub produktem. Wyniki wyszukiwania można również ograniczyć w czasie. NVD wprowadza również możliwość wyszukania wszystkich wpisów CVE, których autorem jest dana jednostka CPE.

Dane znajdujące się w bazie danych NVD przygotowane są do pobrania z oficjalnej strony National Vulnerability Database w formacie XML w wersji 1.2.1, XML w wersji 2.0 lub JSON w wersji testowej 0.1 beta. Do dalszej analizy oraz zastosowania w testach algorytmu został wybrany format XML w wersji 2.0, ponieważ jest to najnowsza, stabilna wersja generowana jako pobieralny plik. W pliku tym znajdują się wszystkie informacje dostępne w ramach bazy danych NVD.

2.3 Analiza pola Opis CVE

Jak wspomniano w sekcji 2.1, pole opisu we wpisach CVE jest jednym z kluczowych pól indeksów CVE, z uwagi na informacje, jakie powinno przechowywać. Szczególnie istotne są te dotyczące produktu, jego wersji oraz dostawcy danego produktu. Niestety brak odpowiedniego, zaplanowanego formatowania utrudnia ekstrakcję tych danych z opisu. Dodatkowo nie zawsze wpisy CVE są w pełni opisane. W wielu przypadkach podatności zgłaszane są przez podmioty, które nie mają dostępu do wszystkich pożądanych informacji. W dodatku liczba osób redagujących opisy sprawia, że pod uwagę należy brać styl pisania, różnice w pisowni lub nazewnictwie produktów, wersji i producentów.