04 ProvChain: oparte na blockchainie potwierdzanie pochodzenia danych w chmurze

Xueping Liang[1], *Sachin S. Shetty*[1], *Deepak Tosh*[2], *Laurent Njilla*[3], *Charles A. Kamhoua*[4] *i Kevin Kwiat*[5]

1 Old Dominion University, Virginia Modeling, Analysis and Simulation Center, Norfolk, VA, USA
2 University of Texas at El Paso, Department of Computer Science, El Paso, TX, USA
3 US Air Force Research Lab, Cyber Assurance Branch, Rome, NY, USA
4 US Army Research Lab, Network Security Branch, Adelphi, MD, USA
5 Haloed Sun TEK, LLC, CAESAR Group, Sarasota, Fl, USA

4.1. Wprowadzenie

Usługi chmury (*cloud computing*) są powszechnie wykorzystywane w środowiskach komercyjnych i wojskowych do realizacji zadań przechowywania danych, udostępniania mocy obliczeniowej na żądanie i dynamicznej alokacji zasobów. Środowiska *cloud computing* są środowiskami rozproszonymi i heterogenicznymi, wykorzystującymi zróżnicowane komponenty programowe i sprzętowe, które mogą pochodzić od wielu dostawców, co może być przyczyną ryzyka lub podatności na ataki bądź niekompatybilności. Kluczowym problemem związanym z zarządzaniem i przesyłaniem danych wewnątrz środowisk chmurowych i między nimi jest zapewnienie bezpieczeństwa. Audyt środowisk chmurowych będzie skuteczny tylko wtedy, gdy można będzie wiarygodnie przeanalizować wszystkie operacje wykonane na danych. Pochodzenie danych (*data provenance*) to proces, który określa historię produktu informacyjnego,

począwszy od jego pierwotnego źródła [1]. Informacje o pochodzeniu danych mogą pomóc w wykrywaniu naruszeń dostępu w infrastrukturach chmury. Dlatego opracowanie metody gwarantowanego potwierdzania pochodzenia danych pozostaje kluczowym problemem w zastosowaniach wykorzystujących chmurę. Informacje o pochodzeniu mogą zawierać wiele poufnych informacji na temat danych źródłowych i ich właścicieli. Dlatego istnieje potrzeba nie tylko zabezpieczenia danych w chmurze, ale także zapewnienia integralności i wiarygodności informacji o ich pochodzeniu. Nowoczesne usługi potwierdzania pochodzenia danych są też podatne na przypadkowe lub złośliwe uszkodzenie bądź próby fałszowania udostępnianych przez nie informacji [2].

Technologia Blockchain rozbudziła zainteresowanie ze względu na oferowaną przez siebie współdzieloną, rozproszoną i odporną na uszkodzenia bazę danych, dzięki której każdy uczestnik sieci może współdziałać na rzecz eliminacji przeciwników, wykorzystując możliwości obliczeniowe uczciwych węzłów. Dzięki temu informacje przekazywane za jej pośrednictwem są odporne na manipulacje. Sieć Blockchain to rozproszony rejestr publiczny, przechowujący transakcje, z których każda jest poświadczona i została zweryfikowana przez węzły sieci.

Zdecentralizowaną architekturę blockchaina można wykorzystać do opracowania mechanizmu dostarczania gwarantowanych informacji o pochodzeniu danych dla środowisk chmurowych. W architekturze zdecentralizowanej każdy węzeł uczestniczy w sieci w celu świadczenia usług, zwiększając w ten sposób jej wydajność. Rozproszony charakter łańcucha bloków poprawia też dostępność. Ponieważ w usługach chmury często wykorzystywany jest jakiś centralny, zaufany podmiot, istnieje potrzeba ochrony wrażliwych danych osobowych przy jednoczesnym zachowaniu prywatności. Oparta na blockchainie usługa informacji o pochodzeniu danych w chmurze pozwoli rejestrować w sposób przejrzysty i trwały wszystkie operacje wykonywane danych. Usługa taka ułatwiłaby budowanie zaufania między użytkownikami a dostawcami usług chmurowych. Ponadto usługa taka mogłaby pomóc w podnoszeniu zaufania użytkowników chmury do udostępnianych informacji o cyberzagrożeniach [3, 4], co zapewniłoby proaktywną cyberobronę przy ograniczonych nakładach na bezpieczeństwo [5, 6].

W tym artykule przedstawiamy ProvChain – opartą na blockchainie architekturę informacji o pochodzeniu danych, udostępniającą pewne informacje na temat operacji na danych w chmurach danych, która jednocześnie zwiększa prywatność i dostępność. ProvChain rejestruje historię operacji jako informacje o pochodzeniu, obliczając ich skróty tworzące węzły drzewa skrótów [7]. Drzewo skrótów zawiera skróty danych o pochodzeniu, a korzeń tego drzewa jest powiązany z jakąś transakcją w łańcuchu bloków. Blok tworzony jest z listy transakcji. Po utworzeniu jest on potwierdzany przez grupę węzłów, co umożliwia włączenie go do łańcucha. Próba zmodyfikowania rekordu

zawierającego informacje o pochodzeniu danych będzie wymagała od przeciwnika zlokalizowania transakcji oraz bloku. Podstawy teoretyczne blockchaina dopuszczają modyfikację rekordu w bloku wyłącznie wtedy, gdy przeciwnik jest w stanie przedstawić dłuższy łańcuch bloków niż łańcuchy pozostałych górników, co jest dość trudne do osiągnięcia. Wykorzystując globalną moc obliczeniową sieci blockchain, oparte na blockchainie informacje o pochodzeniu mogą zapewnić integralność i wiarygodność. W opracowanej przez nas architekturze przechowujemy jedynie skróty danych identyfikujących użytkowników, co chroni ich tożsamość przed innymi węzłami sieci.

Pozostała część artykułu jest uporządkowana następująco: podrozdział 4.2 zawiera przegląd najnowszych metod dostarczania informacji o pochodzeniu danych i omówienie technologii blockchain. Podrozdział 4.3 opisuje projekt architektury ProvChain, czyli zaproponowanej przez nas, opartej na blockchainie, architektury informacji o pochodzeniu danych. Jej szczegółową implementację przedstawiamy w podrozdziale 4.4. Ocenę wydajności architektury ProvChain opisujemy w podrozdziale 4.5. Na koniec, w podrozdziale 4.6 przedstawiamy podsumowanie.

4.2. Kontekst i powiązane prace

4.2.1. Pochodzenie danych

Angielskie słowo „provenance" i jego polski odpowiednik „proweniencja", czyli pochodzenie, wywodzą się z francuskiego słowa „provenir", które znaczy właśnie „pochodzić". Pochodzenie opisuje chronologię zmian własności jakiegoś obiektu. Z punktu widzenia bezpieczeństwa informacji pochodzenie danych oznacza wykaz wszystkich operacji przeprowadzanych na jakichś danych. W kontekście łańcucha bloków, pochodzenie danych można zapisywać w rozproszonym rejestrze publicznym katalogującym wszystkie operacje na danych związanych z jakimś zasobem. Właściciel zasobu może uwierzytelnić transakcję przekazującą prawa własności nowemu właścicielowi bez potrzeby korzystania z niezależnego arbitra. Mechanizm rejestrowania informacji o pochodzeniu danych w łańcuchu bloków może wykorzystywać takie jego właściwości, jak weryfikowalna ścieżka audytu, możliwość tworzenia i przenoszenie praw własności do zasobów cyfrowych, uzgadnianie konsensusu i tożsamości kryptograficzne.