Zarządzanie zasobami danych i ich ochrona

Chris Dotson

Pierwszym krokiem po zapoznaniu się z rozdziałem 1 (WDP 03/2020), w którym omówiono zakresy odpowiedzialności dostawcy i klienta, jest ustalenie, gdzie są lub będą znajdowały się dane oraz to, jak je zabezpieczyć. Często pojawia się wiele nieporozumień dotyczących terminu „zarządzanie zasobami". Czym dokładnie są zasoby i co trzeba zrobić, aby nimi zarządzać? Oczywistą – i nieprzydatną – odpowiedzią jest to, że zasoby są wszystkim, co się posiada. Poniżej przedstawiono więcej szczegółów.

W tej książce zarządzanie zasobami zostało podzielone na dwie części: zarządzanie zasobami danych i zarządzanie zasobami w chmurze. Zasoby danych to ważne posiadane informacje, takie jak nazwy i adresy klientów, dane karty kredytowej, dane rachunku bankowego lub dane uwierzytelniające do uzyskania dostępu do takich danych. Zasoby w chmurze to te elementy, które są wykorzystywane do przechowania i przetwarzania danych: zasoby obliczeniowe, takie jak serwery lub kontenery, pamięć masowa, taka jak magazyny obiektów lub pamięć blokowa, oraz instancje platformy, takie jak bazy danych lub kolejki danych. Zarządzanie tymi zasobami zostało omówione w następnym rozdziale. Możliwe jest rozpoczęcie lektury zarówno od części związanej z zasobami danych, jak i od części związanej z zasobami w chmurze. Aczkolwiek w celu pełnego zrozumienia zagadnień może być konieczne zapoznanie się z informacjami z pozostałych rozdziałów, dlatego też łatwiej jest zacząć lekturę od zasobów danych.

Teoria zarządzania zasobami danych w chmurze nie różni się od tej dla warunków lokalnych, aczkolwiek w praktyce istnieją pewne technologie w chmurze, które mogą być pomocne.

Identyfikacja i klasyfikacja danych

Nawet prosty, omówiony w poprzednim rozdziale schemat i model zagrożenia jest wystarczający, żeby mieć pojęcie o tym, czym są ważne dane, kim są potencjalni atakujący, których trzeba się obawiać, oraz o tym, co chcą oni uzyskać. Poniżej omówiono różne sposoby, w jakie osoby atakujące mogą atakować dane.

Jednym z bardziej popularnych modeli bezpieczeństwa informacji jest triada CIA: poufność, integralność i dostępność. Potencjalni atakujący próbujący naruszyć poufność danych chcą je ukraść, zwykle sprzedać za pieniądze lub też skompromitować właściciela. Potencjalni atakujący próbą naruszenia integralności danych chcą zmienić dane, na przykład przez zmianę salda bankowego. Należy pamiętać, że może to być skuteczne, nawet jeśli osoba atakująca nie może odczytać sald bankowych. Wiele osób z przyjemnością skopiowałoby saldo bankowe rachunku Billa Gatesa, nawet bez znajomości jego aktualnej wartości. Osoba potencjalnie atakująca może próbować naruszyć dostępność danych dla zabawy, zysku lub użyć oprogramowania *ransomware* do zaszyfrowania plików właściciela[1].

Większość z nas ma ograniczone zasoby i musi gospodarować nimi rozsądnie[2]. System klasyfikacji danych może być w tym pomocny, należy jednak oprzeć się pokusie, żeby bardziej go skomplikować, niż jest to absolutnie konieczne.

Przykładowe poziomy klasyfikacji danych

Każda organizacja jest inna, ale poniższe zasady stanowią dobry, prosty punkt wyjścia do oceny wartości posiadanych danych, a zarazem i ryzyka ich naruszenia:

Niska

Informacje z tej kategorii mogą, ale nie muszą, być przeznaczone do publicznego udostępnienia, a gdyby zostały opublikowane, to wpływ na daną organizację byłby bardzo niewielki lub nieistotny. Poniżej przedstawiono kilka przykładów:

- publiczne adresy IP używanych serwerów;
- dane dzienników aplikacji, niezawierające danych osobowych, sekretów ani innych informacji wartościowych dla atakujących;
- materiały instalacyjne oprogramowania bez sekretów lub innych wartości dla atakujących.

Umiarkowana

Informacje te nie powinny być ujawniane poza organizacją bez odpowiednich umów o zachowaniu poufności. W wielu przypadkach, szczególnie w większych organizacjach, tego rodzaju dane powinny być ujawniane wyłącznie w ramach niezbędnej wiedzy w organizacji. W większości organizacji przeważająca część informacji należy właśnie do tej kategorii. Poniżej przedstawiono kilka przykładów:

- szczegółowe informacje na temat projektowania systemów informatycznych, które mogą być przydatne dla osoby atakującej;
- informacje o personelu, które mogą być pomocne dla atakujących w celu przeprowadzenia ataków socjotechnicznych;
- rutynowe informacje finansowe, takie jak zamówienia lub zwrot kosztów podróży, które można wykorzystać na przykład do wywnioskowania, czy przejęcie jest prawdopodobne.

Wysoka

Informacje te są niezbędne dla danej organizacji, a ich ujawnienie może

spowodować znaczną szkodę. Dostęp do tych danych powinien być bardzo ściśle kontrolowany, z wykorzystaniem wielu zabezpieczeń. W niektórych organizacjach tego typu dane nazywane są „klejnotami koronnymi".

Poniżej przedstawiono kilka przykładów:

- informacje o przyszłej strategii lub informacje finansowe, które mogłyby zapewnić znaczącą przewagę konkurentom;
- tajemnice handlowe, takie jak przepis na popularny napój bezalkoholowy lub smażonego kurczaka;
- sekrety zapewniające „klucze do królestwa", takie jak dane uwierzytelniające do pełnego dostępu do infrastruktury chmury;
- wrażliwe informacje przeznaczone do bezpiecznego przechowywania, takie jak dane finansowe klientów;
- wszelkie inne informacje, w przypadku których naruszenie może być warte upublicznienia.

Należy pamiętać, że prawo i reguły branżowe mogą mieć bardzo istotne znaczenie w tym, w jaki sposób klasyfikowane są niektóre informacje. Na przykład Ogólne Rozporządzenie o Ochronie Danych w Unii Europejskiej (RODO) stawia wiele różnych wymagań dotyczących przetwarzania danych osobowych, tak więc w ramach tego rozporządzenia dane osobowe mogą zostać sklasyfikowane jako umiarkowane ryzyko i odpowiednio chronione. Wymagania branży kart płatniczych (PCI) najprawdopodobniej sprawią, że dane posiadaczy kart zostaną sklasyfikowane jako wysokie ryzyko.

Należy również pamiętać, że istnieją usługi w chmurze, które mogą pomóc w klasyfikacji i ochronie danych. Na przykład Amazon Macie (https://amzn.to/2T0ffgA) może być pomocny w znalezieniu poufnych danych w segmentach S3, a interfejs API Google Cloud Data Loss Prevention (http://bit.ly/2GYVoqW) może zostać wykorzystany do klasyfikowania lub maskowania niektórych rodzajów wrażliwych danych.

Bez względu na to, jaki system klasyfikacji zostanie wykorzystany, należy zapisać definicję każdego poziomu klasyfikacji i wybrane przykłady każdego z nich oraz upewnić się, że wszyscy generujący, gromadzący lub chroniący dane rozumieją ten system klasyfikacji.

Istotniejsze wymagania branżowe i prawne

W niniejszej książce poruszono i omówiono sprawy bezpieczeństwa, a nie zgodność z regulacjami branżowymi i prawnymi. Nadmiernie generalizując, zgodność polega na udowodnieniu osobom trzecim, że bezpieczeństwo zostało zapewnione i jest to o wiele łatwiejsze, jeśli systemy i dane zostały faktycznie zabezpieczone. Informacje zawarte w tej książce są pomocne w zachowaniu bezpieczeństwa, niemniej jednak po zabezpieczeniu systemów może być konieczne wykonanie dodatkowych prac związanych z dokumentacją i zapewnieniem zgodności.

Istotne jest, że niektóre wymagania dotyczące zgodności mogą mieć jednak wpływ na projekt zabezpieczeń. Nawet na tym wczesnym etapie ważne jest zwrócenie uwagi na kilka wymagań branżowych lub prawnych:

RODO UE

Niniejsze rozporządzenie może mieć zastosowanie do danych osobowych każdego obywatela Unii Europejskiej lub Europejskiego Obszaru Gospodarczego niezależnie od tego, gdzie na świecie dane są przechowywane. RODO wymaga, aby katalogować, chronić i kontrolować dostęp do „wszelkich informacji odnoszących się do możliwej do zidentyfikowania osoby, którą można zidentyfikować bezpośrednio lub pośrednio, w szczególności przez odniesienie do identyfikatora". Techniki zawarte w tym rozdziale mogą pomóc spełnić niektóre wymagania RODO, ale trzeba się upewnić się, że dołączone są odpowiednie dane osobowe do części danych, które są chronione.

US FISMA lub FedRAMP

Federalna ustawa o zarządzaniu bezpieczeństwem informacji (ang. *Federal Information Security Management Act*) dotyczy poszczególnych agencji, certyfikacja zaś Federalnego Programu Zarządzania Ryzykiem i Autoryzacją (ang. *Federal Risk and Authorization Management Program*) może być stosowana w wielu agencjach, ale obie wymagają klasyfikacji danych i systemów zgodnie z FIPS 199 (http://bit.ly/2BQRBJc) i innymi standardami rządowymi USA. Jeżeli projektowany system wpasowuje się w powyższe rozporządzenia i konieczne jest uzyskanie jednego z tych certyfikatów, powinien zostać wykorzystany poziom klasyfikacji FIPS 199.

ITAR USA

Jeśli firma podlega przepisom dotyczącym międzynarodowego handlu bronią, oprócz własnych kontroli musi wybrać usługi w chmurze obsługującej ITAR. Takie usługi są dostępne u niektórych dostawców usług w chmurze i są zarządzane wyłącznie przez personel z USA.

Globalny PCI DSS

W przypadku firm korzystających z informacji o kartach kredytowych Standard Bezpieczeństwa Danych Kart Płatniczych (ang. *Payment Card Industry Data Security Standard*) narzuca określone mechanizmy kontrolne, które muszą zostać wprowadzone, a także wyszczególnia pewne rodzaje danych, które nie mogą być przechowywane.

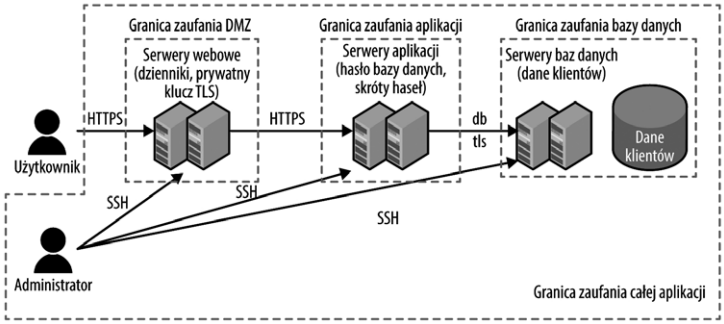
US HIPAA

W przypadku przebywania w USA i zarządzania danymi z chronionymi informacjami zdrowotnymi (PHI) Ustawa o Przenośności i Rozliczalności Ubezpieczeń Zdrowotnych (ang. *Health Insurance Portability and Accountability Act*) nakazuje umieszczanie tych informacji na liście i ich ochronę, co często wiąże się z szyfrowaniem.

Istnieje wiele innych wymagań regulacyjnych i branżowych na całym świecie, takich jak MTCS (Singapur), G-Cloud (Wielka Brytania) i IRAP (Australia). W przypadku podlegania jednemu z nich należy sprawdzić, które rodzaje danych muszą być chronione, aby mieć pewność, że zasoby te zostaną odpowiednio skatalogowane i zabezpieczone.

Zarządzanie zasobami danych w chmurze

Większość omówionych poprzednio zagadnień stanowi dobrą i ogólną



Rys. 1. Przykładowy schemat aplikacji z zasobami danych

praktykę, nie są one jednak specyficzne dla środowiska w chmurze. Dostawcy usług w chmurze znajdują się w wyjątkowej sytuacji, aby pomóc zidentyfikować i sklasyfikować dane. Początkowo będą chcieli, żeby były to wszystkie dane, gdziekolwiek są przechowywane, ponieważ za ich przechowywanie będą mogli naliczyć opłatę!

Ponadto sama specyfika usług w chmurze sprawia, że korzystanie z nich zapewnia już pewien poziom standaryzacji. W wielu przypadkach trwałe dane w chmurze będą znajdować się w jednej z usług, która przechowuje dane, takie jak pamięć obiektowa, pamięć plikowa, pamięć blokowa, baza danych w chmurze lub kolejka komunikatów w chmurze, a nie będą zapisane w rozproszony sposób na tysiącach różnych dysków podłączonych do różnych fizycznych serwerów.

Dostawca usług w chmurze zapewnia narzędzia do inwentaryzacji miejsc przechowywania danych, a także do uzyskiwania do nich dostępu, w starannie kontrolowany sposób, w celu ustalenia, jakie typy danych są tam przechowywane. Dostępne są również usługi w chmurze, które mogą zostać wykorzystane do sprawdzenia wszystkich miejsc przechowywania danych użytkownika i następnie do próby automatycznego sklasyfikowania, gdzie przechowywane są ważne dane. Następnie informacje te mogą zostać wykorzystane do oznaczania zasobów w chmurze, które przechowują dane.

W przypadku identyfikacji ważnych danych nie należy zapominać o hasłach, kluczach API i innych ukrytych informacjach, których można użyć do odczytu lub modyfikacji danych! Dyskusja na temat najlepszych sposobów zabezpieczenia ukrytych informacji została zamieszczona w rozdziale 4, konieczne jest jednak posiadanie wiedzy, gdzie tego typu informacje się znajdują.

W przypadku analizowanej przykładowej aplikacji oczywiste jest, że w bazie danych znajdują się dane klientów. Należy jednak rozważyć, gdzie jeszcze mogą znajdować się ważne zasoby. Poniżej przedstawiono kilka spraw, które powinno się w tym celu rozważyć:

- na serwerach sieciowych przechowywane są dzienniki zawierające dane, które mogą zostać wykorzystane do identyfikacji klientów;
- serwer internetowy wykorzystuje prywatny klucz certyfikatu TLS. Wykorzystując go oraz przejmując niewielką ilość danych DNS lub BGP, każdy może podszywać się pod witrynę i kraść hasła klientów, próbujących się zalogować;
- często prowadzona jest lista skrótów haseł służąca do weryfikacji klientów. Najlepiej jest wykorzystywać wybrany, federacyjny system identyfikacyjny, w przeciwnym razie skróty haseł mogą być dobrym celem[3] dla atakujących;
- serwer aplikacji potrzebuje hasła lub klucza API, aby uzyskać dostęp do bazy danych. Za pomocą tego hasła osoba atakująca może odczytać lub zmodyfikować w bazie danych wszystko to, co może aplikacja.

Nawet w tak prostej aplikacji, jak przedstawiona na rysunku poniżej, jest wiele nieoczywistych rzeczy, które muszą być chronione (rysunek 1).

Oznaczanie zasobów w chmurze

Większość dostawców chmury, a także systemy zarządzania kontenerami, takie jak Kubernetes, oferują koncepcję znaczników. Znacznik jest zwykle kombinacją nazwy (lub „klucza") i wartości. Znaczniki mogą być używane do wielu celów, od kategoryzacji zasobów w spisie komponentów, przez podejmowanie decyzji dotyczących dostępu, kończąc na określaniu zdarzeń, o których należy powiadamiać. Na przykład można mieć klucz danych osobowych (ang. PII, *Personally Identifiable Information*) i wartości *tak* dla wszystkiego, co on zawiera, lub można użyć klucza *typ danych* i wartości klucza danych osobowych.

Problem może stanowić to, że gdyby wszyscy w danej organizacji używali różnych znaczników, to nie byłyby one za bardzo przydatne! Należy więc utworzyć listę znaczników wraz z objaśnieniami, kiedy należy ich używać. Tych samych znaczników powinno się używać u wielu dostawców usług w chmurze, dodatkowo wymagając ich stosowania przez automatyzację (tj. zautomatyzowane narzędzia) podczas tworzenia zasobów. Nawet jeśli jeden z dostawców usług w chmurze nie obsługuje jawnie używania znaczników, to często istnieją inne pola opisu, które mogą służyć do przechowywania znaczników w łatwych do przeanalizowania formatach, takich jak JSON.

Znaczników można używać bezpłatnie, więc ich tworzenie nie jest problematyczne, aczkolwiek dostawcy usług w chmurze mogą nakładać ograniczenia na liczbę znaczników, którą może mieć zasób (zwykle od 15 do 64 znaczników na zasób). Jeśli nie są one potrzebne do kategoryzacji lub podejmowania decyzji, łatwo można je zignorować.

Niektórzy dostawcy usług w chmurze oferują nawet automatyzację w celu sprawdzenia, czy znaczniki są odpowiednio stosowane do zasobów, dzięki czemu można wcześniej wychwycić zasoby nieoznaczone lub błędnie oznaczone i je poprawić. Na przykład, jeśli zdefiniowano zasadę, że każdy zasób musi być oznaczony maksymalną dopuszczalną klasyfikacją danych dla tego zasobu, możliwe jest uruchomienie automatycznego skanowania, tak aby znaleźć wszelkie nieoznaczone zasoby lub takie, dla których wartość znacznika jest inna niż wynikająca z przyjętej klasyfikacji.

Tabela 1. Terminy stosowane do określania oznaczania

Infrastruktura	Nazwa funkcji
Amazon Web Services	Znaczniki
Microsoft Azure	Znaczniki
Google Compute Platform	Etykiety i znaczniki sieciowe
IBM Cloud	Znaczniki
Kubernetes	Etykiety

Mimo że wszyscy główni dostawcy oferują w pewnym stopniu obsługę oznaczania, nie wszyscy zapewniają pełne wykorzystanie możliwości tego typu usługi. Na przykład możliwe jest oznaczanie tworzonych maszyn wirtualnych, ale już nie baz danych. Wszędzie tam, gdzie znaczniki nie są dostępne, konieczne jest klasyczne postępowanie i wykorzystanie ręcznej listy instancji tych usług.

W tabeli 1 zaprezentowano terminologię używaną przez różnych dostawców chmury do określenia oznaczania.

Na obecną chwilę wystarczy zapamiętać niektóre znaczniki, które mogą mieć zastosowanie do różnych zasobów w chmurze, takie jak: *klasa danych: niska; klasa danych: umiarkowana; klasa danych: wysoka; klasa danych: gdpr*.

Ochrona danych w chmurze

Niektóre z technik ochrony danych, które zostały omówione w tej części, można również zastosować w środowiskach lokalnych, aczkolwiek wielu dostawców usług w chmurze zapewnia łatwe, ustandaryzowane i tańsze sposoby ochrony danych.

Tokenizacja

Po co przechowywać dane, kiedy można przechowywać coś, co funkcjonuje podobnie do danych, ale jest bezużyteczne dla osoby atakującej? Tokenizacja, która jest najczęściej używana w przypadku numerów kart kredytowych, polega na zastąpieniu fragmentu wrażliwych danych tokenem, zwykle generowanym losowo. Zaletą jest to, że token ma te same cechy (takie jak długość 16 cyfr) jak oryginalne dane, więc systemy baz danych zbudowane tak, aby przyjmować tego typu dane, nie muszą być specjalnie modyfikowane. Rzeczywiste wrażliwe dane są przechowywane tylko w jednym miejscu, „usłudze tokena". Tokenizacji można używać samodzielnie lub w połączeniu z szyfrowaniem, co omówiono poniżej.

Przykładami usług tokenizacji w chmurze mogą być: usługi współpracujące z przeglądarką w celu tokenizacji poufnych danych przed ich wysłaniem oraz usługi znajdujące się między przeglądarką a aplikacją, które mają na celu tokenizację poufnych danych, zanim te dotrą do aplikacji.

Szyfrowanie

Szyfrowanie jest złotym środkiem w sferze ochrony danych, gdzie pożądane byłoby „zaszyfrowanie wszystkiego". Niestety jest to trochę bardziej skomplikowane. Dane mogą występować w trzech stanach:

- „w ruchu" (przesyłane przez sieć);
- „w użyciu" (obecnie przetwarzane w procesorze komputera lub przechowywane w pamięci RAM);
- „w spoczynku" (w trwałym magazynie, takim jak dysk).

W tym artykule omówiono dwa ostatnie stany, w jakich mogą się znaleźć dane.

Nie zawsze jest wymagana, a nawet przydatna, większa ilość bitów. Na przykład w chwili pisania tego tekstu AES-128 spełnia standardy federalnego rządu USA i jest często szybszy niż AES-256, choć zagrożeniem dla niego mogą być komputery kwantowe. Ponadto algorytm skrótu, taki jak SHA-512, może nie zapewniać żadnej dodatkowej ochrony, jeśli skrót zostanie później przycięty do mniejszej długości.

Szyfrowanie danych podczas ich używania

W momencie pisania niniejszej książki szyfrowanie danych „w użyciu" jest

nowym zagadnieniem i dotyczy przede wszystkim środowisk o bardzo wysokim poziomie bezpieczeństwa. Wymagane jest wsparcie na platformie sprzętowej, które musi zostać ujawnione przez dostawcę chmury. Najczęstsze wdrożenie polega na szyfrowaniu pamięci procesora, tak aby nawet uprzywilejowany użytkownik lub złośliwe oprogramowanie działające jako uprzywilejowany użytkownik nie mogli jej odczytać, a procesor mógł ją odczytać tylko wtedy, gdy ten konkretny proces jest uruchomiony[4]. W przypadku środowisk o bardzo wysokim poziomie bezpieczeństwa, z modelem zagrożeń obejmującym ochronę danych w pamięci przed uprzywilejowanym użytkownikiem, należy skorzystać z platformy obsługującej szyfrowanie pamięci, takich jak: Intel SGX, AMD SME i IBM Z Pervasive Encryption.

Szyfrowanie danych w spoczynku

Prawidłowe wdrożenie szyfrowania danych w spoczynku może być najbardziej skomplikowane. Problemem w tym przypadku nie jest szyfrowanie danych, gdyż istnieje wiele bibliotek pozwalających to zrobić. Problem polega na tym, że po zaszyfrowaniu danych dostępny jest klucz szyfrowania, za pomocą którego można uzyskać do nich dostęp. Niestety wiele osób umieszcza taki klucz tuż obok zaszyfrowanych danych! Przypomina to sytuację, w której zamyka się drzwi, a następnie wiesza klucz na haczyku z etykietą „klucz". W celu zapewnienia prawdziwego bezpieczeństwa (zamiast pozoru zaszyfrowania danych) konieczne jest odpowiednie zarządzanie kluczami. Na szczęście istnieją usługi w chmurze, które są w tym pomocne.

Zaszyfrowanych danych nie można skutecznie skompresować. Jeśli dane mają być skompresowane, trzeba to zrobić przed ich zaszyfrowaniem.

W tradycyjnych środowiskach lokalnych o wysokich wymaganiach bezpieczeństwa można zakupić sprzętowy moduł bezpieczeństwa (HSM, ang. *Hardware Security Module*) do przechowywania kluczy szyfrowania, zwykle w postaci karty rozszerzeń lub modułu dostępnego przez sieć. HSM ma silną logiczną i fizyczną ochronę przed nieautoryzowanym dostępem. W większości systemów każdy, kto ma fizyczny dostęp, może łatwo uzyskać dostęp, aczkolwiek HSM ma czujniki, które usuwają dane, gdy tylko ktoś spróbuje je rozebrać, zeskanować za pomocą promieni rentgenowskich lub dłubać przy źródle zasilania.

Jednakże rozwiązania HSM są drogie i z tego powodu są często niemożliwe do wdrożenia w przypadku większości rozwiązań lokalnych. Jednak w środowisku w chmurze zaawansowane technologie, takie jak HSM i systemy zarządzania kluczami szyfrującymi, są obecnie w zasięgu projektów o skromnym budżecie.

Niektórzy dostawcy usług w chmurze mają możliwość wypożyczenia dedykowanego HSM dla swojego środowiska. Chociaż może to być wymagane w środowiskach o najwyższym poziomie bezpieczeństwa, dedykowany moduł HSM jest nadal drogi w środowisku w chmurze. Inną opcją jest usługa zarządzania kluczami (KMS, ang. *Key Management Service*), usługa wielodostępna, która używa modułu HSM w wewnętrznej bazie danych do zapewnienia bezpieczeństwa kluczy. Konieczne jest więc zaufanie zarówno HSM, jak i KMS (zamiast tylko HSM), co stwarza dodatkowe ryzyko. Jednakże w porównaniu z zarządzaniem tylko własnym kluczem (często niepoprawnie) rozwiązanie KMS zapewnia doskonałe bezpieczeństwo przy zerowym lub bardzo niskim koszcie. Można więc zapewnić korzyści płynące z właściwego zarządzania kluczami w projektach o skromniejszych budżetach.

W tabeli 2 przedstawiono listę kluczowych opcji zarządzania, oferowanych przez głównych dostawców usług w chmurze (do momentu powstania artykułu).

Kolejnym zagadnieniem jest sposób poprawnego skorzystania z usługi zarządzania kluczami. Może to być nieco skomplikowane.

Tabela 2. Opcje zarządzania kluczami

Dostawca	Dedykowana opcja HSM	Usługa zarządzania kluczami
Amazon Web Services	CloudHSM	Amazon KMS
Microsoft Azure	–	Key Vault (klucz programowy)
Google Compute Platform	–	Cloud KMS
IBM Cloud	CloudHSM	Key Protect

Zarządzanie kluczami. Najprostszym podejściem do zarządzania kluczami jest wygenerowanie klucza, zaszyfrowanie danych tym kluczem, umieszczenie klucza w usłudze zarządzania kluczami, a następnie zapisanie zaszyfrowanych danych na dysku wraz z notatką wskazującą, który klucz został użyty do jego zaszyfrowania. Z tym podejściem związane są jednak dwa główne problemy:

1. Zbyt duże obciążenie kiepskich usług zarządzania kluczami. Istnieją dobre powody, dla których pożądane jest posiadanie oddzielnego klucza dla każdego pliku, tak więc usługa zarządzania kluczami z dużą liczbą klientów musiałaby przechowywać miliardy lub tryliony kluczy przy niemal natychmiastowym wyszukiwaniu.
2. W przypadku gdy dane mają zostać usunięte w bezpieczny sposób, konieczne jest upewnienie się, że w usłudze zarządzania kluczami odpowiednie klucze zostaną nieodwołalnie usunięte bez pozostawienia żadnych kopii zapasowych. Można też nadpisać wszystkie zaszyfrowane dane[5], co może być czasochłonne.

Zastąpienie dużej ilości danych może wymagać nie tyle wielu godzin, ile nawet dni. Lepiej jest więc mieć opcję szybkiego i bezpiecznego usuwania obiektów danych na dwa sposoby: przez usunięcie klucza w usłudze zarządzania kluczami, co może skutecznie wymazać wiele różnych obiektów jednocześnie, lub usuwając klucz w miejscu, w którym dane są faktycznie przechowywane, tak aby usunąć pojedynczy obiekt danych. Z tych powodów zazwyczaj istnieją dwa poziomy kluczy: *klucz szyfrowania klucza* i *klucz szyfrowania danych*. Jak sugerują nazwy, *klucz szyfrowania klucza* służy do szyfrowania (lub „spakowania") kluczy szyfrowania danych, które następnie są przechowywane tuż obok danych. Klucz szyfrowania klucza zwykle pozostaje w usłudze zarządzania kluczami i dla bezpieczeństwa nigdy jej nie opuszcza. Spakowane klucze szyfrowania danych są wysyłane do HSM w celu rozpakowania w razie potrzeby, a następnie rozpakowane klucze są używane do szyfrowania lub deszyfrowania danych. Klucze niespakowane nie są nigdy zapisywane. Po zakończeniu bieżącej operacji szyfrowania lub deszyfrowania są kasowane[6].

Użycie kluczy jest łatwiejsze do zrozumienia na przykładzie analogii ze świata rzeczywistego dotyczącej sprzedaży domu. Sprzedawany dom zawiera wszystkie dane. Agent nieruchomości dostaje klucz, aby mógł otworzyć drzwi. Ten domowy klucz jest jak klucz szyfrujący dane i może być wykorzystany do bezpośredniego dostępu do sprzedawanego domu (dane). Pośrednik umieści ten klucz w skrzynce na klucze w drzwiach i zabezpieczy go kodem dostarczonym przez pośrednika w obrocie nieruchomościami. Ten kod jest jak klucz szyfrowania klucza, a usługa pośrednika handlu nieruchomościami, która rozdaje kody, jest jak usługa zarządzania kluczami. W tej lekko naciąganej analogii skrzynka na klucze jest przenoszona do usługi zarządzania kluczami, gdzie wydawana jest kopia klucza z umową zabraniającą wykonania jego kolejnej kopii (zapis na dysk) oraz nakazującą jego wyrzucenie, gdy nie będzie już potrzebny (skasowanie). Tak naprawdę kod otwierający skrzynkę jest zawsze niewidoczny.

W rezultacie, zbliżając się do domu (dane), wiemy, że klucz do danych jest w tym miejscu, ale nie można go otworzyć bez kolejnego klucza lub hasła. Oczywiście w rzeczywistości wystarczyłby młotek i trochę czasu na wyjęcie klucza ze skrzynki. Ewentualnie można też rozbić okno i wtedy żaden klucz nie będzie już potrzebny. Kryptograficznym odpowiednikiem młotka jest odgadnięcie klucza lub hasła użytego do ochrony klucza danych. Zwykle odbywa się to przez wypróbowanie wszystkich możliwości (*brute force*) lub w przypadku haseł – wypróbowanie wielu popularnych haseł („atak słownikowy"). Jeśli algorytm szyfrowania i implementacja tego algorytmu są poprawne, czas oczekiwania, po którym „młotek" dostanie się do skrzynki, jest dłuższy niż czas życia wszechświata.

Szyfrowanie po stronie serwera i klienta. Dobrą wiadomością jest to, że zwykle nie trzeba samodzielnie wykonywać większości operacji związanych z zarządzaniem kluczami! W przypadku większości dostawców usług w chmurze, korzystania z pamięci masowej i usługi zarządzania kluczami oraz włączenia usługi szyfrowania zarządzania kluczami – usługa pamięci masowej automatycznie utworzy klucze szyfrowania danych. Są one później pakowane za pomocą klucza szyfrowania kluczy, którym można zarządzać w usłudze zarządzania kluczami i przechowywać spakowane klucze wraz z danymi. Ciągle można zarządzać kluczami w usłudze zarządzania kluczami i nie jest konieczne samodzielne ich pakowanie lub rozpakowanie, dodatkowo nie trzeba samodzielnie wykonywać operacji szyfrowania i deszyfrowania. Niektórzy dostawcy nazywają to szyfrowaniem po stronie serwera.

Ponieważ usługa pamięci masowej obejmująca wiele podmiotów ma możliwość odszyfrowania danych, błąd w tej usłudze pamięci może potencjalnie umożliwić nieautoryzowanemu użytkownikowi wysłanie żądania do tej usługi o odszyfrowanie danych. Z tego powodu zlecenie usługi przechowywania, szyfrowania i odszyfrowania nie jest tak bezpieczne, jak szyfrowanie we własnym zakresie, oczywiście jeśli zostanie zaimplementowane poprawnie, przy użyciu dobrze znanych bibliotek i procesów. Jest to często nazywane szyfrowaniem po stronie klienta. Aczkolwiek, dopóki nie dotyczy to sytuacji niskiej tolerancji na ryzyko lub budżetu odpowiadającego tej niskiej tolerancji, zaleca

się korzystać z dobrze przetestowanych usług w chmurze i umożliwienie im obsługi szyfrowania i odszyfrowania.

Należy pamiętać, że gdy wykorzystywane jest szyfrowanie po stronie klienta, serwer nie ma możliwości odczytu zaszyfrowanych danych, ponieważ nie ma kluczy. Oznacza to, że nie można wyszukiwać po stronie serwera, obliczać, indeksować, skanować w poszukiwaniu złośliwego oprogramowania ani wykonywać innych zadań o wysokiej wartości. Szyfrowanie homomorficzne może sprawić, że operacje takie, jak dodawanie, będą wykonywane poprawnie na zaszyfrowanych danych bez odszyfrowywania danych, ale na obecną chwilę jest to proces zbyt wolny, aby był praktyczny.

O ile ktoś nie poświęcił większości swojej wybitnej kariery na kryptografię, niezalecane są próby tworzenia i wdrażania własnych systemów kryptograficznych. Nawet podczas samodzielnego szyfrowania i odszyfrowania zalecane jest wykorzystywanie tylko dobrze przetestowanych implementacji bezpiecznych algorytmów, takich jak te zalecane w NIST SP 800-131A (https://bit.ly/2tc1LiC) Rev 1 lub nowszy.

Kasowanie kryptograficzne. Tak naprawdę trudno jest niezawodnie zniszczyć duże ilości danych[7]. Całkowite nadpisanie danych zajmuje dużo czasu, a nawet wtedy mogą znajdować się inne kopie. Możemy to rozwiązać przez wymazanie kryptograficzne. Dzięki takiemu podejściu zamiast przechowywać dane w postaci czystego tekstu na dysku, przechowuje się tylko wersję zaszyfrowaną. Następnie, gdy dane mają zostać niemożliwymi do odzyskania, trzeba wyczyścić lub odwołać dostęp do klucza szyfrowania kluczy w usłudze zarządzania kluczami, co sprawi, że wszystkie klucze szyfrowania danych spakowane w tym kluczu staną się bezużyteczne, gdziekolwiek by się znajdowały na świecie. Możliwe jest również wyczyszczenie konkretnych fragmentów danych, usuwając tylko przypisane im, spakowane klucze szyfrowania danych, dzięki czemu można skutecznie uniemożliwić odzyskanie pliku o objętości wielu terabajtów przez nadpisanie 256-bitowego klucza.

Jak szyfrowanie może blokować różne typy ataków

Jak już wspomniano, szyfrowanie danych w spoczynku może chronić dane przed atakującymi, ograniczając ich możliwości. Dane są wtedy dostępne w sposób bezpośredni tylko w kilku miejscach, w zależności od tego, gdzie odbywa się szyfrowanie. Poniżej przedstawiono typowe udane ataki oraz to, na ile wybory szyfrowania potrafią być utrudnieniem dla atakujących.

Osoba atakująca uzyskuje nieautoryzowany dostęp do fizycznych nośników. Osoba atakująca może z powodzeniem ukraść dyski z centrum danych, śmietnika lub też ukraść taśmy podczas transportu.

Szyfrowanie w spoczynku chroni dane na nośniku fizycznym, dzięki czemu osoba atakująca nie może z nich skorzystać, nawet jeśli uzyska dostęp do nośnika (na przykład przez złamanie hasła). To świetna wiadomość, aczkolwiek ten typ ataku zwykle nie stanowi dużego ryzyka, biorąc pod uwagę fizyczne kontrole i inne zabezpieczenia wdrażane przez większość dostawców chmury. Jest to o wiele ważniejsze w przypadku urządzeń przenośnych, takich jak smartfony i laptopy. Szyfrowanie wykonywane tylko dla zasady często pomaga jedynie złagodzić zagrożenie spowodowane fizyczną kradzieżą, a często nie chroni nawet i przed tym zagrożeniem, ponieważ rozpakowane klucze znajdują się na tym samym nośniku co dane.

Osoba atakująca uzyskuje nieautoryzowany dostęp do platformy lub systemu pamięci masowej. Może zdarzyć się sytuacja, kiedy osoba atakująca lub nieuczciwy dostawca jest w stanie odczytywać i zapisywać dane w bazie danych, blokach pamięci, plikach lub instancji pamięci obiektowej.

Jeśli sam system pamięci masowej jest odpowiedzialny za wykonanie szyfrowania, osoba atakująca jest często w stanie oszukać system tak, aby podał mu dane, w zależności od środków kontroli technicznej w systemie przechowywania. Spowoduje to jednak co najmniej pozostawienie wykrywalnych śladów w zupełnie innym systemie (systemie

zarządzania kluczami), tak więc może być możliwe ograniczenie ataku, jeśli zachowanie podczas dostępu do klucza jest nietypowe i zostanie zauważone przez kogokolwiek wystarczająco szybko.

Jednak w przypadku, gdy aplikacja wysyła do systemu pamięci tylko dane, które są już zaszyfrowane, osoba atakująca uzyska dostęp tylko do bezużytecznego „worka bitów". Osoba atakująca może sprawić, że dane te staną się niedostępne, ale nie może zagrozić ich integralności lub poufności.

Jak wspomniano wcześniej, konieczne jest wyważenie proporcji między zaufaniem, jakim obdarza się system kontroli pamięci masowej, a inwestycją i zaufaniem do własnych środków kontroli. Ogólnie rzecz biorąc, właściciel systemu pamięci masowej ma więcej do stracenia niż użytkownik w przypadku, gdy dojdzie do naruszenia bezpieczeństwa. Sytuacja taka może być szkodliwa dla użytkownika, ale w przypadku dostawcy może to przyczynić się do zakończenia jego działalności.

Atakujący uzyskuje nieautoryzowany dostęp do hypervisora. Większość środowisk w chmurze ma wiele maszyn wirtualnych („gości") działających na hypervisorze, który z kolei działa na fizycznym sprzęcie. Częstym problemem jest to, że osoba atakująca jest w stanie odczytać lub zmodyfikować dane od innych gości na tym samym systemie fizycznym.

Jeśli osoba atakująca jest w stanie odczytać pamięć gościa, może ona także wykorzystać skanowanie pamięci do znalezienia kluczy szyfrujących dane, a następnie użyć ich do odszyfrowania danych. Jest to znacznie trudniejsze niż bezpośrednie odczytywanie danych (a utrudnienie życia atakującego ma wiele zalet). Często jest to jednak możliwe, tak więc jeśli jest to poważne zagrożenie, należy rozważyć wykorzystanie hypervisorcy z pojedynczym klientem, maszyny wirtualnej instalowanej bezpośrednio na sprzęcie lub technologii sprzętowej, która szyfruje dane w pamięci. Jednakże analizując dostępne statystyki dotyczące naruszeń danych, można wywnioskować, że w większości przypadków najprawdopodobniej lepiej jest zainwestować w inne dziedziny związane z bezpieczeństwem.

Osoba atakująca uzyskuje nieautoryzowany dostęp do systemu operacyjnego. Jeśli osoba atakująca uzyska nieautoryzowany dostęp do systemu operacyjnego, na którym działa aplikacja, należy rozważyć dwa scenariusze:

- Osoba atakująca ma ograniczony dostęp do systemu operacyjnego. W tym momencie jedynymi skutecznymi zabezpieczeniami są te należące do systemu operacyjnego. Szyfrowanie w spoczynku nie uniemożliwi dostępu do danych, jeśli osoba atakująca ma dostęp do procesu lub plików przechowujących klucze szyfrujące lub dostęp do odszyfrowanej pamięci.
- Osoba atakująca ma pełny dostęp do systemu operacyjnego. Exploity eskalacji uprawnień są powszechnie dostępne, tak więc osoba atakująca, która zdobędzie ograniczony dostęp do systemu operacyjnego, często bardzo szybko może uzyskać pełne uprawnienia. W przypadku braku omówionych wcześniej zabezpieczeń danych w użyciu osoba atakująca, mając wystarczająco dużo czasu, może odczytać pamięć procesu, odzyskać klucze szyfrowania używane przez wyższe warstwy i uzyskać dostęp do wszystkich danych dostępnych dla tego procesu.

Osoba atakująca uzyskuje nieautoryzowany dostęp do aplikacji. Jeśli osoba atakująca uzyska nieautoryzowany dostęp do aplikacji, wszystko jest przegrane, ponieważ aplikacja musi być w stanie odczytywać dane w celu funkcjonowania. Jednak prawidłowe użycie szyfrowania i innych mechanizmów kontroli dostępu może uniemożliwić osobie atakującej odczytanie danych innych niż dane, do których dostęp ma przejęta aplikacja.

Ogólnie rzecz biorąc, jeśli na spodzie piramidy zostanie umieszczony sprzęt fizyczny, a na szczycie aplikacja, to możliwe jest uzyskanie ochrony przed większą liczbą ataków, umieszczając szyfrowanie tak blisko „wierzchołka" piramidy, jak to tylko możliwe. Kompromis polega na tym, że jest to często bardzo czasochłonne, a należy uwzględnić także prawdopodobieństwo ataku na niższych warstwach.

W wielu przypadkach wkłada się dużo więcej wysiłku w zabezpieczenie niższych warstw niż w zabezpieczenie aplikacji. Dopóki aplikacja nie jest co najmniej tak bezpieczna, jak warstwy pod nią, przeniesienie szyfrowania do samej aplikacji faktycznie zwiększa ryzyko zamiast je zmniejszać. Taki kompromis z aplikacją może zaprzepaścić całe starania poprawy bezpieczeństwa. Z tego powodu zaleca się korzystanie z narzędzi szyfrujących dostępnych w niższych warstwach (zaszyfrowane bazy danych, szyfrowanie bloków i plików itp.) w przypadku większości projektów. Szyfrowanie na poziomie aplikacji zalecane jest tylko dla bardzo wrażliwych danych, ze względu na to, że przy znacznie większej czasochłonności osiągane jest jedynie minimalne zmniejszenie ryzyka.

Podsumowanie

Przy planowaniu strategii środowiska w chmurze konieczne jest określenie typu posiadanych danych, zarówno ich części oczywistej, jak i nieoczywistej. Należy sklasyfikować każdy typ danych według negatywnych konsekwencji w przypadku, gdyby dane zostały odczytane, zmodyfikowane lub usunięte przez osobę atakującą. Po uzgodnieniu w całej organizacji, które znaczniki mają być używane w „słowniku znaczników", należy wykorzystać funkcje oznaczania, które oferowane są przez dostawcę usług w chmurze do oznaczenia zasobów zawierających dane.

Jeśli tylko jest to możliwe, to przed utworzeniem instancji pamięci należy zdecydować się na strategię szyfrowania, ponieważ późniejsza zmiana może być trudna. W większości przypadków należy używać systemu zarządzania kluczami szyfrowania dostawcy usług w chmurze, a także szyfrowania wbudowanego w usługi pamięci masowej, jeśli są dostępne, akceptując ryzyko naruszenia bezpieczeństwa usługi pamięci masowej. Jeżeli konieczne jest samodzielne zaszyfrowanie danych przed ich zapisaniem, należy korzystać tylko z dobrze przetestowanych i bezpiecznych algorytmów.

Konieczne jest ostrożne kontrolowanie użytkowników i systemów, które mają dostęp do kluczy, oraz takie skonfigurowanie powiadomień, aby informowały, kiedy klucze są uzyskiwane w nietypowy sposób. Zapewnia to kolejną, oprócz kontroli dostępu w instancjach pamięci, warstwę ochrony, a także umożliwia łatwy sposób na kryptograficzne usunięcie informacji, gdy nie są już potrzebne.

Jedną z częstych obaw związanych z szyfrowaniem jest to, że może zmniejszyć wydajność ze względu na dodatkowy czas przetwarzania wymagany do zaszyfrowania i odszyfrowania danych. Na szczęście nie jest to już tak duży problem, jak kiedyś. Sprzęt jest tani, a wszyscy główni producenci układów oferują wbudowane przyspieszenie sprzętowe w swoich procesorach. Problemy z wydajnością rzadko są dobrym pretekstem do nieszyfrowania danych, aczkolwiek zupełną pewność można uzyskać tylko dzięki rzeczywistym testom.

Ważniejszym problemem związanym z szyfrowaniem jest dostępność danych. Jeśli nie można uzyskać dostępu do kluczy szyfrowania, nie można uzyskać dostępu do danych. Należy zapewnić sobie dodatkowy proces („wejście awaryjne") pozwalający na uzyskanie dostępu do kluczy szyfrowania z jednoczesnym sprawdzeniem, czy proces ten nie może być wykorzystany bez wyraźnego sygnalizowania i ostrzegania.

Przypisy

1. *Ransomware* stanowi zarówno naruszenie dostępności, jak i integralności, ponieważ wykorzystuje nieautoryzowane modyfikacje danych w celu uczynienia ich niedostępnymi.
2. Jeśli masz nieograniczone zasoby, skontaktuj się ze mną!
3. Można wspomnieć 6,5 miliona skrótów LinkedIn, które zostały złamane, a następnie wykorzystane do naruszenia bezpieczeństwa innych kont, których użytkownicy używali tego samego hasła co na LinkedIn.
4. Należy pamiętać, że szyfrowanie w pamięci chroni dane tylko przed atakami spoza procesu. Jeśli uda się atakującemu wykorzystać ten proces do zrobienia czegoś, czego nie powinien, może on odczytać pamięć i dane.
5. Pomimo wniosków ze znanego dokumentu USENIX z 1996 r. (http://bit.ly/2U4QRXK), gdzie badano możliwość odzyskania danych z dysku twardego, który został nadpisany, dziś nie jest to praktyczne. Odzyskiwanie zastąpionych danych z dysków półprzewodnikowych (SSD) jest nieco bardziej praktyczne ze względu na sposób zapisu, ale większość dysków SSD ma funkcję „bezpiecznego wymazywania" w celu bezpiecznego wyczyszczenia całego dysku. Więcej szczegółów można znaleźć w dokumencie USENIX z 2011 r. Michaela Wei i in. (http://bit.ly/2Vj7SxO).
6. To jest bardzo uproszczone wyjaśnienie. Naprawdę głęboką dyskusję na temat wszystkich zagadnień kryptograficznych można znaleźć w książce Bruce'a Schneiera *Applied Cryptography* (Wiley).
7. Chociaż, paradoksalnie, często łatwo to zrobić przez przypadek!

Fragment pochodzi z książki: *Bezpieczeństwo w chmurze. Przewodnik po projektowaniu i wdrażaniu zabezpieczeń*,
Chris Dotson, Wydawnictwo Naukowe PWN, Warszawa 2020