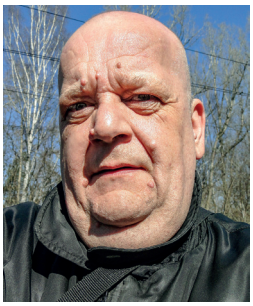
Dane kruche jak porcelana

Żyjemy w świecie Big Data. „Wielka dana" to nasz bożek, fetysz i nadzieja na lepsze jutro. Dzięki danetyzacji świata, czyli kwantyfikowaniu i przetwarzaniu wszelkich aspektów ludzkiego życia do postaci cyfrowych danych komputerowych, możemy obecnie podziwiać „cuda" sztucznej inteligencji, która bez olbrzymiej ilości wprowadzonej informacji byłaby tylko pustym algorytmem. Potrzeba kompulsywnego zwiększania ilości danych jest więc zrozumiała, a każda utrata cennych danych jest pod wieloma względami niepowetowana.



Jacek Grabowski

z wykształcenia specjalista gazownictwa i górnictwa naftowego, przygodę z informatyką rozpoczął w końcu lat 80. XX wieku od współpracy z wydawnictwem „Lupus", gdzie publikował teksty głównie w dwutygodniku „PCkurier" i miesięczniku „Enter". Współtwórca pierwszego w Polsce informatycznego czasopisma B2B „MRK" (1997). Były redaktor naczelny miesięcznika „Reset", współpracownik wielu innych tytułów (magazyn „WWW", „IT Reseller", „Komputer Świat"). Obecnie freelancer, współpracuje m.in. z warszawską komunikacją miejską.

Straty danych mają wymiar finansowy – oblicza się, że firmy tracą średnio milion dolarów w roku z powodu wycieków lub innej formy utraty danych. Utrata danych może mieć także wymiar osobisty – wolumen danych tworzonych wewnętrznie przez prywatnych użytkowników jest również olbrzymi, a obejmują one najintymniejsze sfery życia, są też często cenną pamiątką. Trudno się dziwić, że kwestia możliwości odzyskania utraconych danych jest z roku na rok ważniejsza, a know-how w tym względzie wysoko cenione na rynku.

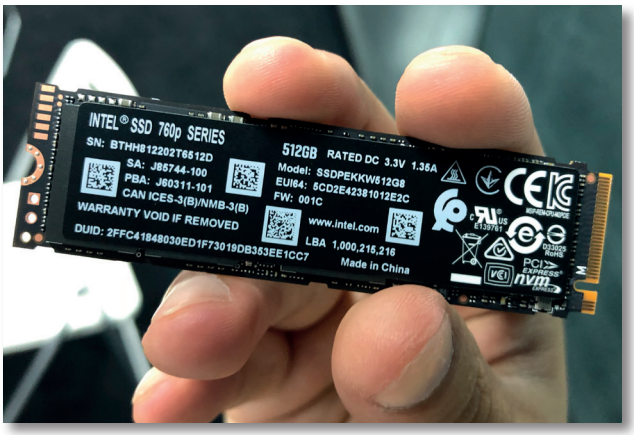
Jak przechowujemy dane?

Żeby dane stracić, trzeba je najpierw pozyskać, a następnie po odpowiednim przetworzeniu zapisać na właściwym nośniku. Od początku szerszej komputeryzacji powszechnie stosowano zapis magnetyczny danych: najpierw analogowym sygnałem na taśmie magnetycznej, takiej jak w zwykłych magnetofonach, potem już cyfrowo na magnetycznych dyskach. Potocznie nośniki danych są określane jako „dyski", co utrwaliło także wprowadzenie nośników wykorzystujących zapis optyczny sygnału cyfrowego, czyli płyt CD i DVD. Jednak w ostatnich czasach pojęcie „dysku" zaczęło zatracać pierwotny sens, a to za sprawą rozwoju nośników opartych na pamięciach flash. Dlatego zgrabniej jest używać słowa „napęd", chociaż w przypadku SSD również i to słowo stanowi przenośnię.

Wnętrze napędu HDD, widać talerz magnetyczny i głowicę na ramieniu.

Najbardziej rozpowszechnione nośniki danych, napędy twardych dysków określane też jako HDD (*Hard Disk Drive*), zapisują dane na wirujących dyskach magnetycznych. Takich dysków (talerzy) w jednym napędzie może być kilka, zwykle są dwa do czterech. Operacje zapisu/odczytu na dyskach są dokonywane przez umieszczone na przesuwnych ramionach głowice. Przypomina to trochę klasyczny gramofon na płyty winylowe, przy czym głowice nie dotykają powierzchni dysku i przesuwają się znacznie bardziej precyzyjnie, o setne części milimetra, trafiając dokładnie w miejsca, które wskaże układ kontrolera współpracujący z systemem operacyjnym. Nośnik tego typu ma wiele elementów mechanicznych, które ograniczają jego niezawodność i prędkość działania. Jego wyrafinowana konstrukcja wymaga niezwykłej precyzji montażu w sterylnie czystych pomieszczeniach. Obecnie napędy HDD są najczęściej używane do przechowywania dużych ilości danych przy komputerach stacjonarnych.

Napędy optyczne, czyli odczytujące/zapisujące płyty CD/DVD, są stosowane w naszych czasach bardzo rzadko i duża część komputerów osobistych nie jest w ogóle w nie wyposażona. Z kolei napędy HDD są obecnie masowo zastępowane przez taniejące SSD (Solid State Drives, „dyski" czy też napędy półprzewodnikowe), w których dane są zapisywane i przechowywane w układach pamięci nieulotnej flash. Takie rozwiązanie pozwoliło uniknąć zastosowania kosztownej, skomplikowanej i zawodnej mechaniki, zastępując wszystkie ruchome części elektroniką. Dzięki temu wzrosła między innymi prędkość transmisji danych, umożliwiając opracowanie nowych interfejsów bezpośrednio kontaktujących się z magistralą danych CPU.



Intel 512 SSD w formacie M.2 przyjmuje postać wąskiej płytki, niekiedy obudowanej radiatorem odprowadzającym ciepło.

Napędy SSD są dzisiaj stosowane np. w telefonach, notebookach i tabletach. Służą jako dyski systemowe i wewnętrzna pamięć. Najczęściej przyjmują formę płytki z interfejsem M.2 wpinanej bezpośrednio do płyty głównej komputera. Istnieją także napędy SSD z interfejsem SATA pasujące do miejsca po 2,5-calowym HDD, które można instalować np. w starszych notebookach bez złącza M.2. Warto pamiętać, że popularne „pendrajwy", czyli nośniki wkładane bezpośrednio w złącze USB, mimo że również wykorzystują pamięci flash, zasadniczo różnią się od SSD. Stosowane są w nich gorsze (wolniejsze) rodzaje pamięci, a także odmienne kontrolery, które inaczej obsługują operacje zapisu/odczytu.

Dane i metadane

Mimo istnienia różnych rozwiązań zapisu danych na nośnikach, dla komputera jest obojętne, jakiego rodzaju napędu użyjemy do przechowania informacji. Wystarczy, że system operacyjny ma odpowiedni program obsługi, który kontaktuje się z kontrolerem danego urządzenia. Jednak każdy system operacyjny ma swój system plikowy, więc żeby mógł obsługiwać zapis i odczyt na dysku wytworzonych przez nas danych, najpierw musi na nim zapisać tzw. metadane, czyli logiczną strukturę systemu plików. Stąd bierze się dalsze rozróżnienie nośników fizycznych od „dysków" logicznych, utworzonych przez system operacyjny.

Dyski logiczne to tzw. partycje, czyli wirtualne struktury metadanych rozpoznawane przez komputer jako osobny dysk fizyczny. Na jednym rzeczywistym nośniku może być jedna albo nawet kilka partycji. Po utworzeniu partycji jest ona następnie formatowana, czyli otrzymuje kolejne metadane opisujące szczegółowo system plików i umożliwiające ich identyfikację oraz położenie na dysku. Istnieją też metadane niższego poziomu, niewidoczne dla systemu operacyjnego i użytkownika, związane z firmware'em i kontrolerem napędu.

Dopiero po zdefiniowaniu struktury nośnika i podzieleniu dysku na sektory, cylindry czy klastry można zacząć zapisywanie danych wytwarzanych w różnych programach. Dane te są umieszczane w plikach, czyli jak gdyby pojemnikach nadających im odpowiedni format i porządek. Rozróżniamy przy tym kilka rodzajów plików, np.: wykonywalne zawierające programy, multimedialne np. z muzyką albo filmami, graficzne z obrazami, itd. Z punktu widzenia systemu operacyjnego są to tylko zbiory liczb, dla nas – litery, cyfry, zdjęcia.

Logiczna i sprzętowa utrata danych

Rozróżnienie między danymi a metadanymi ma duże znaczenie w procesie odzyskiwania danych. Jeżeli uszkodzona jest logiczna struktura dysku, to istnieje duże prawdopodobieństwo, że dane zapisane na tak zniszczonym nośniku uda się odzyskać w całości za pomocą odpowiednich programów, które potrafią odtworzyć logiczną strukturę dysku. Jeżeli zniszczony jest dysk fizyczny, czyli urządzenie, to przynajmniej część danych może okazać się nie do odzyskania, a samo odzyskanie będzie wymagało znacznie bardziej skomplikowanych zabiegów niż uruchomienie programu.

Do zniszczenia logicznej struktury dysku może dojść na skutek działania zewnętrznych czynników, na przykład malware, czyli złośliwego oprogramowania celowo modyfikującego lub niszczącego metadane nośnika np. dla wymuszenia okupu w celu odzyskania danych. Ale to nie wszystko. Dane może przecież omyłkowo zlikwidować sam użytkownik, wydając niewłaściwe polecenie skasowania pliku czy sformatowania od nowa dysku twardego. Systemy operacyjne są „idiotoodporne" i utrudniają użytkownikom wykonanie błędnych operacji na zapisanych danych. Służy temu np. typowy dla wszystkich współczesnych systemów dwufazowy proces kasowania pliku: najpierw przesunięcie do wirtualnego „kosza", z którego można plik odzyskać, a potem dopiero ostateczna likwidacja. Podobnie utrudnione jest sformatowanie dysku systemowego, ale już np. dysk zewnętrzny można sformatować dużo łatwiej. W dodatku pliki z pendrive'a nie przechodzą weryfikacji w koszu, a są kasowane po prostym potwierdzeniu decyzji. Tak więc pomyłki mimo wszystko mogą się zdarzać. Do uszkodzenia metadanych mogą prowadzić także np. nieprzewidziane wyłączenia komputera (np. utrata zasilania).

Fizyczne uszkodzenia nośnika też mogą przybierać różne formy – niekiedy przyczyniają się do nich użytkownicy, np. upuszczając dysk na ziemię z pewnej wysokości, zalewając komputer kawą lub wodą czy też łamiąc kartę pamięci albo pendrive'a. Spora część takich uszkodzeń skutkujących utratą dostępu do danych wynika z awarii elektroniki (kontrolera) nośnika. Awarie takie często wynikają po prostu ze starzenia się układów i użytkownik nie musi się do tego przyczynić.

Utrata danych czy utrata dostępu?

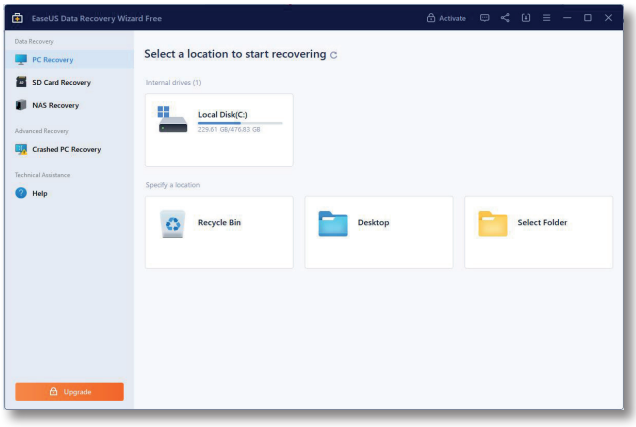
W większości przypadków jakiekolwiek uszkodzenie struktury danych czy fizycznego nośnika na szczęście nie prowadzi od razu do bezpowrotnej utraty danych. Dane te pozostają gdzieś zapisane, tylko tracimy do nich dostęp. Komputer zgłasza taki nośnik jako niedziałający albo w ogóle go nie rozpoznaje, jednak niemożność odczytania jego zawartości nie oznacza, że jej tam nie ma. Giną dopiero dane celowo wymazane fizycznie bądź nadpisane na dysku inną zawartością.

Dyski magnetyczne charakteryzują się pewną cechą – gdy kasujemy plik, to nie zostaje on fizycznie wymazany, tylko system operacyjny zaznacza sobie, że został zlikwidowany. Dopiero gdy zachodzi potrzeba zapisania nowych danych w miejscu skasowanego pliku poprzednie dane zostają bezpowrotnie zniszczone poprzez nadpisanie. Również gdy omyłkowo sformatujemy dysk, stare dane nie są od razu zamazywane, czyszczona jest tylko tablica informująca, gdzie były położone na dysku. Żeby zamazać dane podczas formatowania trzeba użyć specjalnej procedury formatowania niskopoziomowego zapisującego wszystkie sektory zerami – dopiero wówczas dane są praktycznie nie do odzyskania.

Gorzej wygląda sytuacja danych zapisanych na nośniku typu SSD. Tam niestety stare dane muszą zostać najpierw wymazane, żeby można było zapisać nowe. Gdyby jednak takie wymazywanie działało w czasie rzeczywistym, to spadałaby prędkość zapisu. Żeby tego uniknąć, producenci SSD zalecają stosowanie specjalnego polecenia systemowego TRIM, które informuje napęd, że skasowane dane nie są potrzebne. Wtedy kontroler napędu przesuwa sektory z tymi danymi do puli przeznaczonej do późniejszego wymazania i traktuje je tak, jakby były zapisane śmieciami, tzn. nie ma sposobu odczytania znajdujących się w nich danych. Dlatego odzyskiwanie danych z napędów SSD jest dużo trudniejsze niż z HDD, a praktycznie niemożliwe po ich omyłkowym sformatowaniu lub usunięciu plików. Po uszkodzeniu logicznej struktury odzyskanie danych jest możliwe, lecz co najmniej wymaga wyłączenia polecenia TRIM.

Programowe odzyskiwanie danych

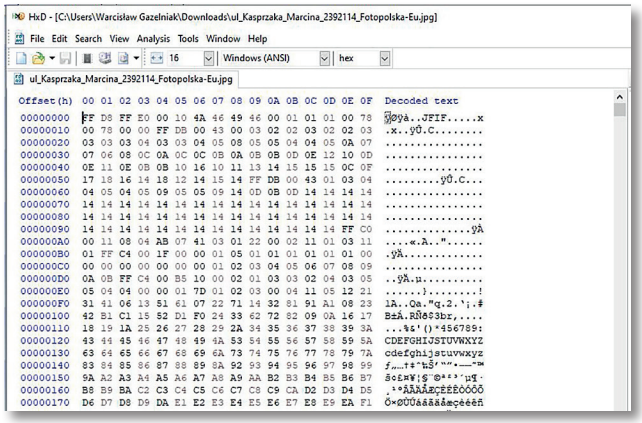
W przypadku uszkodzenia metadanych na dysku lub błędnej ingerencji użytkownika w system plików (sformatowanie lub usunięcie danych) pomocny przy odzyskiwaniu będzie specjalizowany program. Takich programów istnieje bardzo dużo na rynku, ale charakteryzują się one wspólnymi cechami. Muszą umożliwić dostęp nawet do napędu, który jest sprawny, ale przestał być normalnie rozpoznawany przez system operacyjny. Po uzyskaniu dostępu do dysku wskazanego przez użytkownika jako uszkodzony, program taki rozpoczyna analizowanie systemu plików. Po znalezieniu i odczytaniu odpowiednich metadanych, na ich podstawie jest odbudowywana struktura plików i katalogów. Przy mniejszym zakresie uszkodzeń możliwe jest natychmiastowe odzyskanie przez program wszystkich plików.



Ekran startowy programu Easeus Data Recovery Wizard, bezpłatna wersja pozwala odzyskać 512 MB danych.

Tak korzystna sytuacja występuje jednak rzadko. Zwykle program nie jest w stanie automatycznie odtworzyć wszystkich danych z uszkodzonego napędu i część z nich wykazuje jako „dodatkowe pliki" w folderach z automatycznie nadanymi

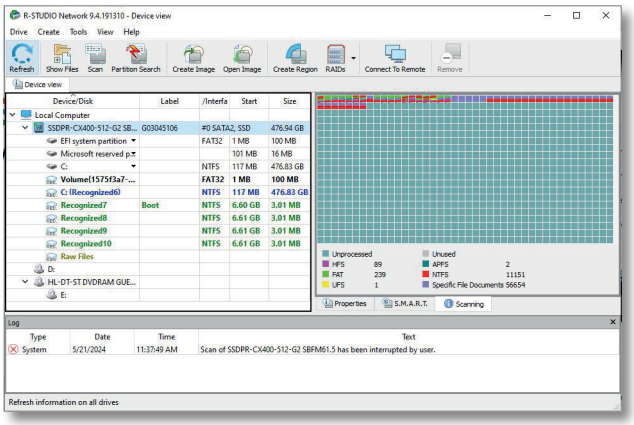
tymczasowymi nazwami. Wtedy zaczyna się żmudniejsza praca odzyskiwania tzw. danych surowych (*raw*), czyli de facto binarnej magmy znajdującej się w przestrzeni dysku, której struktury logicznej nie dało się w pełni odtworzyć. Program analizuje tę magmę pod kątem zawartości, szukając np. znaczników początku i końca pliku. Np. pliki JPG z obrazami mają na początku sygnaturę JFIF, a wykonywalne pliki .EXE systemu DOS i Windows charakterystyczny znak MZ, będący notabene inicjałami programisty Marka Zbikowskiego zaangażowanego w tworzenie systemu DOS.



Ekran edytora zawartości dysku z szesnastkową reprezentacją danych pliku graficznego w formacie JPG, widoczna jest „rozszyfrowana" zawartość nagłówka z sygnaturą JFIF pozwalającą rozpoznać rodzaj pliku.

Odzyskiwanie danych surowych nigdy nie może się udać w pełni automatycznie. Wpływ na to ma np. fakt, że istnieją pliki bez rozpoznawalnych sygnatur początku i końca, a poza tym na dyskach w mniejszym lub większym stopniu występuje tzw. fragmentacja, czyli zapisywanie jednego pliku w kawałkach umieszczanych w oddalonych od siebie sektorach dysku. Dlatego programy do odzyskiwania danych poza automatycznymi procedurami udostępniają także szereg zaawansowanych narzędzi służących do penetrowania zawartości dysku i wyszukiwania zapisanych na nim informacji. Specjalista odnajduje pewne ciągi znaków lub cyfr i stara się stopniowo odtworzyć wszystkie połączenia między nimi, niemalże budując ręcznie pliki od nowa. Ta żmudna i wymagająca cierpliwości oraz wiedzy praca to rzeczywisty proces „odzyskiwania" danych bajt po bajcie.

Tak więc nie łudźmy się – automatycznie możemy odzyskać swoje dane tylko w bardzo korzystnych sytuacjach, gdy uszkodzenia są niewielkie. Jeśli omyłkowo skasujemy jeden czy dwa pliki, to automat odzyska je bez problemu, o ile nie zostały nadpisane i nie znajdowały się na dysku SSD. Jednak kiedy posypała się cała struktura dysku, to wszystkie automatyczne „wizardy" mogą więcej zaszkodzić, niż pomóc. Dlatego nie mając odpowiedniej wiedzy, lepiej ich używać z umiarem lub nie używać wcale.



Ekran profesjonalnego programu do odzyskiwania danych R-Studio Network z widocznym wynikiem częściowego skanowania dysku w celu zlokalizowania zagubionych danych.

Fizyczna rekonstrukcja dysku

Znacznie gorzej jest, kiedy nośnik danych zostaje uszkodzony fizycznie. Dajmy na to – padnie elektronika sterująca napędem HDD. W takiej sytuacji wydaje się, że wyjściem jest zdobycie nowej płyty głównej dysku i zamontowanie jej w miejsce uszkodzonej. Jednak wcale nie jest to takie proste i oczywiste. Producenci dysków nie sprzedają części zamiennych, poza tym często wprowadzają w trakcie produkcji zmiany technologiczne, więc musi być to dokładnie taka sama wersja płyty, pasująca idealnie do właściwej serii produkcyjnej. W dodatku awaria elektroniki mogła skutkować dalszymi uszkodzeniami już we wnętrzu samego napędu, a wtedy wymiana płyty głównej nic nie da.

W istocie więc trzeba właściwie ocenić stopień uszkodzenia napędu, a w razie uszkodzenia mechaniki ingerować w jego wnętrze. W tym celu należy mieć odpowiednie warunki pracy, zapewniające możliwość otworzenia hermetycznie zamkniętej „bańki" zawierającej talerze i mechanikę napędu, a także precyzyjne narzędzia i odpowiednią wiedzę. Koszty odzyskania danych rosną więc znacznie w porównaniu z przypadkami awarii struktury logicznej.

Jak widzimy z tego krótkiego przeglądu, utrata danych może nam grozić w każdej chwili i nie jest to żart ani straszenie. Wprawdzie istnieją systemy typu S.M.A.R.T. monitorujące stan dysku i uprzedzające o awariach, ale nie we wszystkich przypadkach działają tak, jak należy. Zwłaszcza napędy SSD często przestają działać bez żadnego uprzedzenia. Dlatego o dane należy dbać, robić kopie zapasowe i starać się zawsze być z nimi na bieżąco. A jak już coś się zdarzy i nie wiemy, co to spowodowało – lepiej skorzystać z pomocy specjalisty niż z internetowych porad.