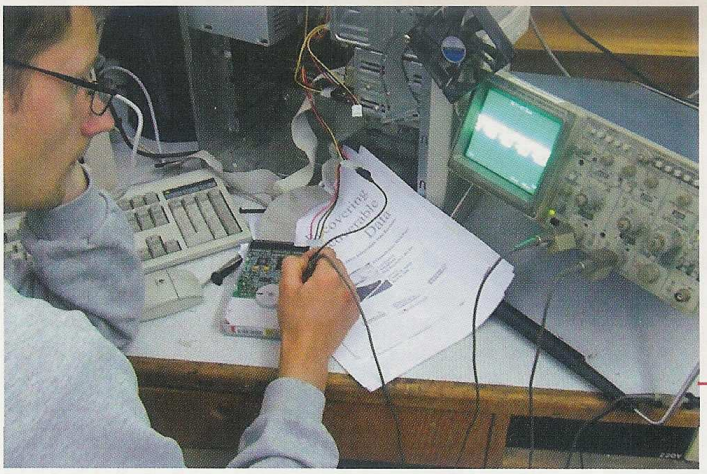
Dane nie do utracenia

BADANIA | Dane, które przechowujemy na dysku, bywają bardziej niebezpieczne, niż nam się wydaje. Naukowcy z PAN relacjonują dla nas projekt badawczy, w którym studenci bez trudności zakupili dużą liczbę „twardzieli", a następnie odzyskali zapisane na nich informacje. Co znaleźli?

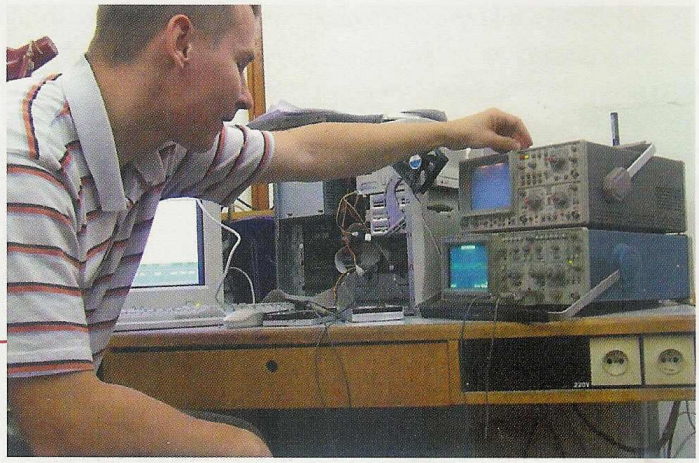
Szymon Piłat, Kamil Kulesza

Coraz częściej docierają do nas informacje o zdarzeniach związanych z niezachowaniem staranności przy przechowywaniu i przetwarzaniu danych elektronicznych. W polskich mediach głośno było choćby o dysku twardym komputera Aleksandry Jakubowskiej, na którym znaleziono obciążające ją dowody, czy „twardzielach" pochodzących z Ministerstwa Spraw Zagranicznych zakupionych przez tygodnik „NIE". Takie przypadki są nie tylko polską domeną – na całym świecie wiele jest podobnych zdarzeń.

Czy chronimy swoje dane w dostatecznym stopniu? Postanowiliśmy to zbadać. Interesowało nas także to, w jakim stopniu Polacy są świadomi nieskuteczności kasowania danych. Chcieliśmy zwrócić uwagę na to, że informacje, które przechowujemy na dysku, bywają często bardziej niebezpieczne, niż nam się wydaje. Projekt badawczy, który przeprowadziliśmy w Polskiej Akademii Nauk, polegał na zakupie jak największej liczby używanych dysków twardych, a następnie na sprawdzeniu, czy znajdują się na nich jakieś istotne dane. Na tej podstawie chcieliśmy opracować skuteczne metody odzyskiwania danych oraz ich analizy.



Jeden ze studentów **podczas pracy z uszkodzonym dyskiem twardym.** Analiza „twardzieli" była dużo większym wyzwaniem niż sam zakup starych dysków (nośniki kupowane były głównie na aukcjach internetowych).



Do eksperymentów z **niskopoziomowym odzyskiwaniem danych** wykorzystano oscyloskopy o wysokich częstościach próbkowania.

Wystarczy Internet

Uczestnikami projektu byli studenci warszawskich uczelni. Co istotne, przeprowadzony projekt miał bardzo niski budżet – zaledwie kilkaset złotych. Było to jednak celowe, nie wynikało z jakichś zewnętrznych ograniczeń. Wiadomo przecież, że dysponując dużą kwotą można zdobyć wszystko, nawet poufne dane z dysków twardych. To jednak byłoby zbyt łatwe. Co warte podkreślenia – przez całe badania udało się nie przekroczyć zaplanowanego budżetu.

Innym ważnym warunkiem było to, że nie zapoznaliśmy studentów z żadnymi zaawansowanych technologiami związanymi z odzyskiwaniem danych. Całą wiedzę czerpali głównie z Internetu. Oprogramowanie, za którego pomocą odzyskiwali zawartość dysków, jest ogólnodostępne w Sieci.

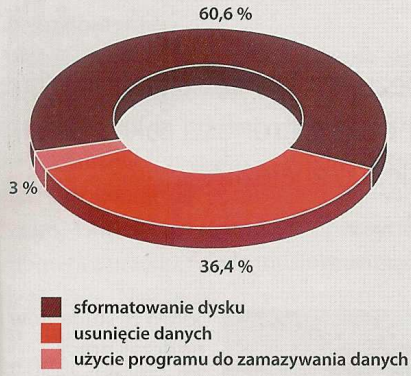
Dzięki takim założeniom pokazaliśmy, że skuteczne odzyskiwanie danych na dużą skalę wymaga jedynie minimalnych nakładów oraz wiedzy i kompetencji na poziomie studenta nauk ścisłych/technicznych.

Miasto pełne dysków

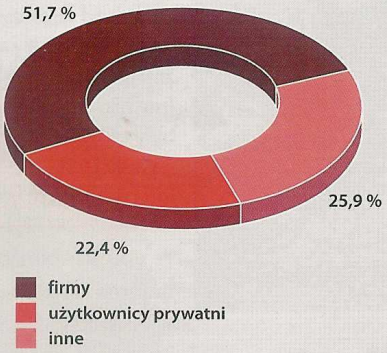
Pierwszym krokiem było zaopatrzenie się w nośniki – dyski twarde studenci kupowali głównie na aukcjach internetowych. Następnie badaliśmy, jakie dane da się z nich odzyskać, po czym „twardziele" odsprzedawaliśmy. Dzięki temu zbadanie jednego dysku kosztowało kilka złotych lub też wręcz „nic". Odnaleźliśmy również kilka miejsc „na mieście", w których można było kupić tanio stare nośniki. Korzystając z tych sposobów, w ciągu trzech miesięcy pozyskaliśmy około 200 dysków. Badanie dysków wskazało, że najczęściej były one sformatowane lub też dane były po prostu usunięte, tak więc ich >>

FOT. P. SPRYSZAK

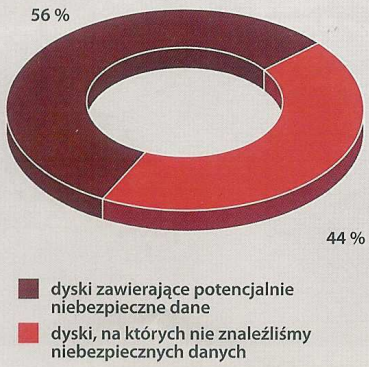
WYNIKI ANALIZY DYSKÓW PRZEBADANYCH PRZEZ STUDENTÓW



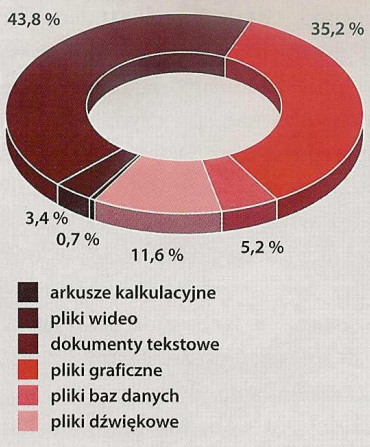
Najpopularniejszą metodą „kasowania" danych było formatowanie. Jedynie 3% dysków zostało zamazanych programem do kasowania danych.



Najwięcej dysków pochodziło od firm i urzędów (prawie 52%). Co czwarty „twardziel" należał wcześniej do użytkownika domowego. 22% nośników pochodziło ze źródeł, których nie dało się sklasyfikować lub ich ustalenie było niemożliwe (tak było np. w przypadku dysków prawidłowo wyczyszczonych).



Na 56% nośników znajdowały się **informacje potencjalnie niebezpieczne**. Pozostałe 44% to dyski zawierające wyłącznie system operacyjny lub prawidłowo wyczyszczone.



Największy odsetek danych potencjalnie niebezpiecznych (43,8%) zawierały dokumenty tekstowe, niewiele mniej takich treści znaleziono w plikach graficznych (35,2%). Najmniej „groźnych" informacji było w arkuszach kalkulacyjnych (3,4%).

odzyskanie nie sprawiało kłopotu. Było to jednak zadanie czasochłonne.

Największym wyzwaniem była analiza zawartości nośników. Gigabajty danych zawierały pliki systemowe i dźwiękowe, filmy, bezużyteczne zdjęcia... Jedynie niewielka część informacji możliwych do odzyskania była istotna z naszego punktu widzenia. Oddzielenie ich od reszty odbywało się częściowo automatycznie, ale ostateczną analizę trzeba było wykonać ręcznie. Niezbędne było dokładne opisanie zawartości dysków, dlatego właśnie w projekcie brało udział kilkanaście osób, dzięki którym udało się przeanalizować w kilka tygodni tak wiele treści.

Najciekawsze znaleziska

Pierwsze wnioski są dość zaskakujące: tylko 3% dysków wyczyszczono należycie, a z pozostałych 97% prawie zawsze udało się odzyskać jakieś dane. Najczęściej były to informacje o posiadaczu komputera, o tym, czym zajmował się w pracy i w życiu prywatnym. Ponad połowa nośników pochodziła z firm i urzędów. Zapisane na nich dane były dużo bardziej „interesujące" od tych, które zawierały „twardziele" należące do użytkowników prywatnych. Udało się na przykład dotrzeć do szczegółowych informacji finansowych jednego ze znanych warszawskich hoteli. Wśród nich znajdowały się pliki z zestawieniami inwestycji, zobowiązań, przepływów pieniężnych, inwentaryzacji, wykaz majątku, opis wartości posiadanych gruntów, a także wszystkie faktury i listy płac. Mogliśmy dowiedzieć się m.in., że osoba zajmująca się sprawami finansowymi, korzystając ze służbowego komputera, napisała do urzędu gminy... donos na swoich sąsiadów z bloku.

Równie ciekawe treści zawierała więcej niż połowa nośników. Okazało się, że niemal każdy użytkownik wykonywał w godzinach pracy czynności niezwiązane z powierzonymi mu obowiązkami, co tylko zwiększało „atrakcyjność" odzyskanych danych. Dysk należący niegdyś do jednostki państwowej pełen był ofert przetargowych. Na nośniku tym (oraz na mniej więcej co czwartym urzędowym i firmowym „twardzielu") znajdowało się rozliczenie podatkowe (najczęściej więcej niż jedno). Około 10% dysków firmowych zawierało treści pornograficzne w postaci klipów wideo, a na dużej pamięci cache

przeglądarki zapisane były głównie zdjęcia w podobnej „tematyce".

Interesującym znaleziskiem były dwa dyski pochodzące z urzędu gminy, na których odnaleźliśmy kilkadziesiąt tysięcy dokumentów: podań, wypowiedzeń, pism związanych z wynajmem lokali oraz kompletną bazę zarejestrowanych działalności gospodarczych na terenie gminy wraz z adresami i telefonami.

Wśród dysków należących wcześniej do firm prywatnych najciekawszy był egzemplarz zawierający dane dotyczące jednej z dużych sieci supermarketów spożywczych. Dowiedzieliśmy się z niego, jakie towary (ilości, ceny i kody) zamawiane były przez tę sieć na terenie całej Polski. Informacje zostały przejrzyście przedstawione w postaci dużej bazy danych oraz gotowych raportów.

Nośniki po reanimacji

Warto podkreślić, że wśród badanych przez nas dysków były również egzemplarze niesprawne. Wydaje się, że ich właściciele nader lekkomyślnie podeszli do tematu bezpieczeństwa danych. Być może uznali, że skoro dysk nie działa, to dane zapisane na nim przepadły. Nie jest to prawda, ponieważ niejednokrotnie „naprawa" takiego nośnika sprowadzała się do wymiany elektroniki bądź głowic czytających – rzadko musieliśmy przeprowadzać bardziej skomplikowane operacje.

Jedną z osób, które pozbyły się takiego „nieczytelnego" dysku, był kierownik firmy budowlanej pracującej przy przebudowie jednego z wieżowców w centrum Warszawy. Po wymianie głowic w nośniku pochodzącym ze służbowego komputera udało się ustalić, że właściciel przechowywał na nim wszystko, co mogłoby być potrzebne potencjalnemu szantażyście. Oprócz informacji na temat pracy (szczegółów dotyczących projektu, kosztorysów inwestycji itp.) na dysku znajdowała się korespondencja kierownika z żoną i... kochanką z Lublina. Wraz ze szczegółowymi danymi, a nawet zdjęciami.

Opisaliśmy tu zaledwie kilka ciekawych przykładów zawartości dysków. Jednak takie przypadki stanowią aż 56% zawartości urzędowych i firmowych nośników, które poddaliśmy analizie! Dane z pozostałych 44% dysków zostały w odpowiedni sposób skasowane lub nie można było ich uznać za niebezpieczne (np. pozostałość zawartości nośnika stanowił sam system operacyjny czy kilka programów).

Krążące dane

Jedną z idei naszego projektu było sprawdzenie, czy głośne afery związane z bezpieczeństwem danych wpływają na świadomość użytkowników. Ponadto chcieliśmy dowiedzieć się, czy informacje pozostawione na dyskach twardych mogłyby potencjalnie zaszkodzić osobom, których dotyczą.

Uzyskane wyniki pokazały jednoznacznie, że ani przeciętny użytkownik domowy, ani firmy czy urzędy nie dbają należycie o bezpieczeństwo swoich danych. Stwarza to okazje do szantaży, kradzieży tożsamości i innych podobnych nadużyć.

Istnieje również inna hipoteza tłumacząca lekkomyślność użytkownika sprzedającego swój dysk: być może zdaje on sobie sprawę z tego, że dane można odzyskać, ale wydaje mu się, że nikt nie zada sobie tyle trudu. Zapewne myśli, że wszystko, co mogłoby być poufne, zostało skasowane, i przekazując dysk nowemu właścicielowi, jest przekonany, że nic szczególnego już na nim nie ma. Tymczasem przy tak dużej pojemności, którą cechują się produkowane dziś „twardziele", nie potrzebujemy zbyt często kasować danych. Często więc okazuje się, że poufne informacje wciąż są na nośniku – tylko ktoś o nich zapomniał.

Lekceważenie kwestii bezpieczeństwa danych może przysporzyć wiele kłopotów. Skuteczne niszczenie danych z dysku twardego nie jest trudne (o tym, jak zrobić to skutecznie, czytaj na str. 20). Mamy do dyspozycji wiele programów kasujących dane, są wśród nich także darmowe. Warto zatem z nich skorzystać, zanim powierzymy swój dysk osobie trzeciej. Nigdy nie jesteśmy w stanie przewidzieć, jakie będą dalsze jego losy. Może trafi do kolejnego projektu badawczego prowadzonego przez naszą grupę? To byłaby wersja z happy endem. Może być jednak tak, że dysk dostanie się w ręce kogoś, kto tylko szuka sposobności do wykorzystania zapisanych na nim danych.

Mówi się, że użytkownicy komputerów dzielą się na tych, którzy robią backupy, i tych, którzy będą je robić. Można dodać, że również na tych, którzy kasują dane z dysków, i na tych... którzy dopiero będą kasować. ir

Szymon Piłat jest studentem Wydziału Fizyki Uniwersytetu Warszawskiego. Od kilku lat zajmuje się m.in. analizą sygnałów oraz problematyką bezpieczeństwa i odzyskiwania danych. W opisywanym w artykule projekcie pełnił funkcję lidera grupy.

Kamil Kulesza jest adiunktem w Instytucie Badań Systemowych PAN oraz pracownikiem naukowym University of Cambridge. W swojej pracy badawczej zajmuje się zagadnieniami związanymi z informatyką i zastosowaniami matematyki, ze szczególnym uwzględnieniem bezpieczeństwa informacji. Jest też zaangażowany w Letnie Praktyki Badawcze PAN od momentu ich powstania.

Kradzież i zgubienie danych

Informacje zapisane na dyskach mogą znaleźć się w „tarapatach" także w sytuacji, gdy zostanie skradziony komputer. Taki przypadek odnotowano w czerwcu w Wielkiej Brytanii – ze szpitala zostało skradzionych sześć laptopów, zawierających informacje o około 20 000 pacjentach. Trzy miesiące później, także w UK, ujawniono skandal związany ze zgubieniem pendrive'a przez pracownika firmy pracującej na zlecenie rządu brytyjskiego. Na nośniku były dane dotyczące 10 000 przestępców i 84 000 więźniów.

Takim przypadkom można zaradzić nie tyle przez kasowanie danych, ile ich szyfrowanie oraz zachowanie odpowiednich procedur przechowywania i przenoszenia.

Więcej o projekcie badawczym

Opisywane w artykule badania przeprowadzone zostały w ramach letniej praktyki badawczej (**www.praktyki.ibspan.waw.pl**), prowadzonej przez Polską Akademię Nauk. Projekty, w których uczestniczą studenci zorganizowane są na wzór brytyjskich *study groups*. Tematyka badań dotyczy głównie bieżących problemów przemysłowych i biznesowych, których rozwiązaniem może być podejście badawczo-naukowe. Praktyki są prowadzone w PAN od kilku lat, obecnie są organizowane wspólnie przez Instytut Badań Systemowych PAN, Instytut Matematyczny PAN i Wyższą Szkołę Informatyki Stosowanej i Zarządzania pod auspicjami PAN.