Gazeta Wyborcza, 2016-05-12

Inwigilacja w warunkach terroryzmu i gigadanych

Wojciech Cellary

W toczącej się dyskusji nad projektem zmiany ustawy o policji pojawiają się argumenty za i przeciw, jednak dość powierzchowne i niesięgające do sedna sprawy, a na pewno niepokazujące związków z nowoczesnymi technikami informatycznymi, takimi jak analiza gigadanych (ang. big data).

Państwo a ochrona obywateli

U podstaw problemu leży zjawisko terroryzmu, w szczególności ataków samobójczych. Dotychczasowy system prawny opierał się na następujących założeniach: Państwo przekazywało społeczeństwu przekaz: obywatelu nie popełniaj przestępstwa, bo policja cię złapie, prokurator oskarży, sędzia osądzi, a za karę twoje życie będzie skrócone, bo zamknięcie w więzieniu to jak skrócenie życia. Jednak ten przekaz nie działa w odniesieniu do terrorystów samobójców, bo oni popełniając zbrodnię zabicia przypadkowych, niewinnych ludzi sami wymierzają sobie najwyższy wymiar kary – śmierć. Bardziej ukarać ich nie można. Obowiązkiem Państwa jest chronić swoich obywateli, ale w tym przypadku jedynym sposobem ochrony jest niedopuszczenie do ataku na drodze aresztowania osób, które taki atak planują. To jednak oznacza konieczność naruszania ich prywatności, aby dowiedzieć się o ich planach z wyprzedzeniem. Stąd bierze się zmiana podejścia Państwa (a raczej wielu Państw zagrożonych terroryzmem) do inwigilacji obywateli. Dawniej inwigilacja, w szczególności elektroniczna, była możliwa tylko w odniesieniu do konkretnych obywateli, co do których istniało uzasadnione podejrzenie o popełnienie przestępstwa, i tylko za zgodą sądu stojącego na straży wolności obywatelskich i sprawującego nadzór nad organami ścigania. W epoce terroryzmu Państwo chce zbierać wszelkie możliwe dane o wszystkich obywatelach, czyli wszystkich inwigilować, i w razie potrzeby przeprowadzać analizę zebranych danych w odniesieniu do podejrzanych. Przy takim podejściu człowiek jest traktowany jako suma swoich relacji społecznych, interakcji elektronicznych i ulubionych treści. Obywatel staje się podejrzany nie dlatego, że zrobił coś złego, tylko dlatego, że jego relacje z innymi ludźmi i treściami wskazują na większą niż u innych skłonność do popełnienia przestępstwa.

Gigadane

Takie masowe zbieranie danych – mówimy w tym przypadku o gigadanych (ang. big data) – nie byłoby uzasadnione, gdyby nie istniały skuteczne metody ich analizy. Analiza gigadanych jest dziedziną badań, w której wielka skala opisu zjawisk masowych jest drogą do odkrycia nowej

wiedzy. Tej wiedzy nie można odkryć w małej skali. Ale ta wiedza ma inny charakter, opiera się bowiem nie na poszukiwaniu przyczynowości, tylko korelacji. To zdecydowanie nie jest to samo, ale taka wiedza również może być podstawą skutecznego działania. Tu jednak kryje się niebezpieczeństwo, bowiem dzięki analizie gigadanych będziemy wiedzieli co się dzieje, a nawet – z wysokim prawdopodobieństwem – co się stanie, ale nie będziemy wiedzieli dlaczego.

Przykładowo, na podstawie analizy gigadanych możemy wykryć, że branie dwóch leków jednoczenie zwiększa skuteczność leczenia. Nie musimy wiedzieć dlaczego tak się dzieje, aby podawać pacjentom oba leki. Analiza gigadanych nie ma na celu nauczenia komputerów, aby myślały jak ludzie, tylko zastosowaniu matematyki do wyliczenia prawdopodobieństwa wystąpienia pewnego faktu. Na przykład, po napisaniu jakiegoś słowa w SMS-ie na smartfonie, jakie słowo z największym prawdopodobieństwem będzie następne – to słowo system mi podpowie. Przy czym ta analiza jest spersonalizowana i tym analiza gigadanych różni się od profilowania. Jeśli ja pisząc SMS na smartfonie postawię kropkę, to system zaproponuje mi jako następne słowo „Wojtek", bo statystycznie ja po kropce najczęściej podpisuję się swoim imieniem. Ale komuś innemu zaproponuje takie słowo, które w odniesieniu do niego ma najwyższe prawdopodobieństwo. Jeśli ja napiszę pierwsze dwie litery „po" to system zaproponuje mi słowo „Poznań", ale komuś innemu zaproponuje słowo „Pozdrowienia". Analiza gigadanych jest czymś innym niż profilowanie. Profilowanie polega na kwalifikacji pojedynczego człowieka do określonej grupy – np. mężczyzn powyżej 50 roku życia, albo kierowców poniżej 25 roku życia i zastosowaniu do nich jednakowych reguł, co jest formą odpowiedzialności grupowej – na przykład uznanie ich za zagrożonych zawałem serca, lub powodujących liczne wypadki drogowe. Prawda jest taka, że nie wszyscy mężczyźni po 50 roku życia są zagrożeni zawałem i nie wszyscy młodzi ludzie jeżdżą jak wariaci, więc na podstawie analizy gigadanych, a nie profilowania, tylko niektóry zostaną zakwalifikowani do zagrożonych zawałem i stanowiących zagrożenie na drodze. Odwracając tę sytuację, pasażer o arabskim nazwisku i wyglądzie, płacący gotówką za bilet lotniczy nie musi być automatycznie podejrzany o terroryzm, jeśli z analizy gigadanych dotyczących jego osoby to nie wynika.

Jednak odkrycie korelacji na podstawie analizy gigadanych to nie okrycie związków przyczynowo-skutkowych. Można wykryć nieoczywistą korelację między bezawaryjnością samochodów, a ich pomarańczowym kolorem. Jednak pomalowanie konkretnego samochodu na pomarańczowo, nie spowoduje, że nie będzie się psuć.

Ochrona prywatności

Na proste pytanie – dlaczego chronić prywatność? – jest prosta odpowiedź – ponieważ utrata prywatności prowadzi do podatności na manipulację i dyskryminację. Niestety analiza gigadanych nie tylko zwiększa zagrożenie utraty prywatności, ale zmienia charakter zagrożeń. Wartość gigadanych jest bowiem związana z ich wtórnym, a nie tylko pierwotnym użyciem, i to do celu innego niż cel pierwotny. Podważona jest zatem obecna zasada jawnego wyrażania

zgody przez każdego obywatela na użycie jego danych osobowych do określonego celu i zakazu używania tych danych do wszelkich innych celów. Nie można z góry wyrazić zgody na sposób wykorzystania danych, który jeszcze nie istnieje. Nie można prosić powtórnie właścicieli danych osobowych – liczonych często w setkach milionów ludzi, bo chodzi o gigadane – o zgodę na ich nowe wykorzystanie, gdy taki wtórny cel się pojawi. Jednak nawet brak zgody nie jest zabezpieczeniem. Ludzie protestowali przeciwko pokazywaniu ich domów w usłudze Google Street View w obawie przez złodziejami, jednak rozmazanie zdjęcia konkretnego domu też może być wskazówką dla złodziei. Anonimizacja, czyli usuwanie danych osobowych, umiarkowanie skuteczna w przypadku małych zbiorów danych jest nieskuteczna w przypadku gigadanych, w szczególności pochodzących z łączonych źródeł. Przeprowadzony eksperyment wykazał, że oceny wystawione anonimowo dla 6 z 500 filmów w Netflixie pozwalają zidentyfikować klienta z 84% dokładnością, a jeśli dodatkowo znany jest dzień wystawienia oceny, to z 99% dokładnością.

Sprawiedliwość

Predykcyjna analiza gigadanych, pozwalająca z dużym prawdopodobieństwem przewidzieć z góry co się stanie, umożliwia zmianę podejścia Państwa do ochrony obywateli przed zagrożeniami typu terroryzm, bo pozwala aresztować terrorystów zanim przeprowadzą atak. Jednak takie podejście stanowi zagrożenie dla sprawiedliwości takiej, jaką uznajemy w krajach demokratycznych. Myśl o popełnieniu przestępstwa nie jest bowiem nielegalna – ilu z nas w przypływie emocji chciało „zabić" swojego szefa? Dopiero przejście od takiej myśli do czynu jest nielegalne. Osobista odpowiedzialność jest związana z indywidualnym wyborem. Złodziej, który włamał się do sejfu, zostanie pociągnięty do odpowiedzialności, ale osoba, która otworzyła sejf, ponieważ złodziej przyłożył jej pistolet do głowy, nie będzie pociągnięta do odpowiedzialności. Uznanie kogoś za winnego przewidywanych czynów, których jeszcze nie popełnił, to błąd polegający na wykorzystaniu prognoz z analizy gigadanych bazujących na korelacji do podejmowania decyzji o czyjejś indywidualnej odpowiedzialności, która wymaga istnienia związku przyczynowo-skutkowego: popełnione przestępstwo – kara. Analiza gigadanych pozwala poznać ryzyko wystąpienia przyszłych zjawisk i odpowiednio dostosować do niego nasze działania. Jednak analiza gigadanych nie mówi nic o przyczynowości. Każda osoba musi indywidualnie podjąć decyzję o swoim konkretnym działaniu, które spowoduje konkretny skutek, za który będzie ponosić odpowiedzialność, w tym odpowiedzialność karną. Istnieje niebezpieczeństwo nadużycia analizy gigadanych do szukania przyczynowości na podstawie korelacji, co jest błędem metodologicznym. Nadużycie analizy gigadanych prowadzi do społeczeństwa, w którym nie istnieje wolna wola i indywidualny wybór człowieka, moralny kompas każdego człowieka jest zastąpiony algorytmem prognostycznym, a jednostki są poddane nieograniczonemu przymusowi kolektywnych decyzji. Innymi słowy Państwo ogłasza obywatelowi – my Państwo (policjant, prokurator, sędzia) wiemy, że postąpisz nie według własnego uznania (moralności) tylko tak, jak to wynika z zachowań kolektywnych. Ponieważ to,

co możesz zrobić jest złe, to zamkniemy cię prewencyjnie w więzieniu za przestępstwo, którego wprawdzie nie popełniłeś, ale które możesz popełnić, bo podobni do ciebie je popełnili. Takie podejście jest niedopuszczalne na gruncie sprawiedliwości w państwach demokratycznych, bo oznacza w istocie zniewolenie społeczeństwa. Pomimo możliwości, jakie oferuje predykcyjna analiza gigadanych, ludzie powinni być (nadal) sądzeni za to, co faktycznie zrobili w przeszłości, a nie za statystyczną prognozę tego, co mogą zrobić w przyszłości.

Konflikt wartości

Zbierać, czy nie zbierać gigadanych o obywatelach? Nie ma łatwej odpowiedzi na to pytanie, bo niezależnie od zastosowanego rozwiązania zawsze ktoś może ucierpieć. Albo niektórzy obywatele ucierpią od ataku terrorystycznego, który nie został wyprzedająco wykryty i zneutralizowany przez służby specjalne niemające dostępu do gigadanych, albo niektórzy obywatele ucierpią, bo skorumpowany urzędnik mający dostęp do gigadanych wystawił ich gangsterom do ataku kryminalnego lub nieetyczny polityk będący u władzy mający dostęp do gigadanych dyskryminuje ich za poglądy. Wybór jak w starożytnej greckiej tragedii – konflikt podstawowych wartości.

Jednak przez atakami terrorystycznymi nie ma innej, skutecznej ochrony, a przed skorumpowanymi urzędnikami i nieetycznymi politykami – są. Analiza gigadanych jest następnym krokiem technologicznym ludzkości i jak każda nowa technologia może być wykorzystana do czynienia dobra i zła. Dobro wynikające z analizy gigadanych polega na lepszym przewidywaniu przyszłości i dzięki niemu skuteczniejszej i wcześniejszej ochronie ludzi przed zagrożeniami – terroryzmem, katastrofami, wypadkami, chorobami itp. – oraz lepszym zgadywaniu ich życzeń. Jednak dostęp do gigadanych ludzi złej woli może doprowadzić do ogromnych nadużyć. Te same techniki, które są rozwijane w intencji typowania potencjalnych terrorystów, mogą być użyte do typowania celów ataków kryminalnych, albo dyskryminacji przeciwników politycznych. Dlatego analiza gigadanych musi pozostawać pod skuteczną kontrolą takich organów jak niezawisłe sądy, które muszą mieć do swojej dyspozycji kompetentnych informatyków, bo żaden prawnik tego nie skontroluje. Mamy w Polsce Inspektora Ochrony Danych Osobowych, chroniącego obywateli przed nieuprawnionym przetwarzaniem ich danych osobowych, ale będziemy potrzebować Inspektora Analiz Gigadanych chroniącego obywateli przed nieuprawnionymi wnioskami z takich analiz.

Na zakończenie zachęcam Czytelników do przeczytania książki Kennetha Cukiera i Viktora Mayera-Schönbergera pt. „Big Data. Rewolucja, która zmieni nasze myślenie, pracę i życie”, MT Biznes sp. z o.o., Warszawa 2014, na podstawie której częściowo powstał ten artykuł.

Prof. dr hab. inż. Wojciech Cellary jest kierownikiem Katedry Technologii Informatycznych Uniwersytetu Ekonomicznego w Poznaniu.