Warszawa, 08.08.2024 r.

Czego uczą nas „brzegowe przypadki”?

Na pewno doskonale znany programistom termin „edge case” (z ang.: „*przypadek brzegowy*”) oznacza sytuacje co prawda teoretycznie możliwą, ale której wystąpienie w realnym świecie jawi się jako niezwykle mało prawdopodobne. Zazwyczaj ich pojawienie wiąże się z nietypową sytuacją, w przypadku algorytmów – wprowadzeniem bardzo dużej bądź bardzo małej wartości liczbowej (stąd „brzeg” w nazwie). Warto pamiętać o takich sytuacjach również na gruncie ochrony danych osobowych, co opiszę poniżej na kilku ciekawych przykładach. Przykładem takiej sytuacji może być tak zwana „pluskwa roku 2000” (Y2K bug) – czyli sytuacje polegające na tym, że autorzy programów komputerowych pisanych w latach siedemdziesiątych i osiemdziesiątych nie przewidzieli, że będą one działały przez 20-30 lat i zakodowane obowiązkowe formatowanie bieżącej daty zaczynające się od 19.. może stanowić problem.

Zgodnie z zasadą „taniej uczyć się na cudzych błędach” zapraszam na subiektywny przegląd przypadków brzegowych, które ziściły się w realnym świecie.

1. „My zawsze odbieramy uprawnienia do wszystkich systemów IT przy rozwiązaniu umowy o pracę”

To zdanie, które wypowiedziane z pełnym przekonaniem często słyszę w trakcie audytów. Zawsze pytam wtedy, czy jest stosowana lista kontrolna przy offboardingu – bo to dokonała metoda, żeby upewnić się czy wszystko zostało zrobione jak trzeba za każdym razem. Nie wiemy, czy tego typu zabezpieczenie stosował jeden z dużych polskich banków. To co wiemy – z komunikatu Prezesa Urzędu Ochrony Danych Osobowych[1] i z decyzji nakładającej na ów bank karę w wysokości blisko 550.000 zł [2]– to to, że po zakończeniu współpracy z osobą posiadającą dostęp do Platformy Usług Elektronicznych Zakładu Ubezpieczeń Społecznych tego dostępu jej nie odebrano. Skutek? Osoba ta dysponowała nieuprawnionym, nieograniczonym dostępem do wrażliwych danych osobowych około 10.000 pracowników banku przez okres niecałego roku od zakończenia pracy w tym podmiocie.

Co gorsza, systemy informatyczne ZUS nie rejestrowały kto i do jakich danych uzyskuje dostęp, zatem dokładne ustalenie kręgu osób pokrzywdzonych tą sytuacją okazało się niezwykle trudne. Bo skoro doszło do naruszenia, to trzeba powiadomić osoby, których dane dotyczą. I jest to obowiązek administratora danych osobowych.

Listy kontrolne to świetne rozwiązanie – szczególnie wtedy, gdy określony zestaw czynności jest realizowany rutynowo, w dziesiątkach czy setkach kolejnych iteracji.

2. Te darmowe narzędzia internetowe do generowania albo tłumaczenia tekstów są super, a „za darmo” to dobra cena

Fakt istnienia narzędzi pozwalających wykonywać błyskawiczne tłumaczenia tekstów na obce języki lub generować całe teksty na zadany temat skutecznie przebił się do powszechnej świadomości pracowników. W niedawnym badaniu czytamy, że aż 87% Polek korzysta z narzędzi SI (przy średniej europejskiej 68% [3]). Truizmem będzie powiedzieć, że takie narzędzia pozwalają znacznie

przyspieszyć realizację różnych zadań. Ale korzystając z tych zdobyczy nowoczesnej technologii nie wolno zapominać o tym, jak te systemy działają. A działają w taki sposób, że wszystko co do nich wprowadzimy może pozostać w tych systemach na długo (jak długo, zależy oczywiście od danej usługi i jej polityki prywatności). Oznacza to, że firmy powinny zachować szczególną ostrożność przy wprowadzaniu do takich narzędzi danych osobowych, czy informacji, które objęte są klauzulami poufności.

Dlaczego?

Po pierwsze, nie spotkałem się jeszcze z klauzulą informacyjną z której by wynikało, że moje dane mogą być przetwarzane przy użyciu zewnętrznych narzędzi opartych na uczeniu maszynowym. Formalnie administrator powinien informować osoby, których dane dotyczą o tym, co z tymi danymi będzie się dziać. Coraz głośniej też mówi się o procesach[4] przeciwko dostawcom tych narzędzi o bezprawne korzystanie przez nie z cudzych danych osobowych.

Po drugie, musiałby być niesamowity zbieg okoliczności, żeby dane które wprowadziliśmy do okienka Chat GPT wyciekły, prawda? Jaka jest szansa, że hakerzy przejmą dane do logowania (a więc dostęp do pełnej historii zapytań do algorytmu) ponad 100.000 podmiotów, w tym 1158 kont wykorzystywanych przez podmioty z siedzibą w Polsce? Po ilości szczegółów w poprzednim zdaniu na pewno domyśliliście się Państwo, że taki „przypadek brzegowy" już wystąpił – w 2023 roku[5].

Reasumując: w celu uniknięcia kar za naruszenie RODO bądź umów o zachowaniu poufności – warto z rozwagą korzystać z tego typu narzędzi. Dane osobowe w tekście do tłumaczenia zawsze można zastąpić „X" lub innym znakiem (czyli, jak to mówi RODO – dokonać pseudonimizacji).

3. Jak często trzeba resetować samolot Boeing 787 Dreamliner?

Z pewnych przyczyn technicznych niektóre komputery liczą ilość sekund (a czasami setnych sekund), która upłynęła od ich ostatniego uruchomienia. Programiści mogą przeznaczyć na ten cel dedykowaną ilość pamięci – np. 32 bity. 32 bity to ilość pamięci, która pozwala zapisać 2.147.483.648 liczb całkowitych. To strasznie dużo, prawda? I chyba nie potrzeba przeznaczać na to więcej zasobów, bo raz na jakiś czas komputer będzie resetowany, więc licznik ponownie ruszy od zera.

Jak się jednak okazało, zapisanie po kolei 2.147.483.648 liczb, z których każda odpowiada jednej setnej sekundy, zajmie mniej więcej 248 dni [6]. Po upływie tego czasu (o ile wcześniej system nie zostanie zresetowany), systemowi skończy się dedykowane miejsce do zapisywania tego, od jak dawna jest włączony, więc... prewencyjnie przejdzie w tryb awaryjny[7]. I wyłączy całą elektrykę w samolocie. Problem ten jest na szczęście znany już od 2015 roku i dawno został naprawiony, ale to dobry przykład „przypadku brzegowego", który może mieć bardzo duże konsekwencje.

O podobnej sytuacji niedoszacowania słyszałem na odcinku ochrony danych osobowych. Była sobie firma, która napisała program, który robił „coś". Co dokładnie nie jest istotne, ale gdy firma zaczynała działalność, spodziewano się kilkunastu, kilkudziesięciu klientów. By nie przeciążać bazy danych podjęto decyzję, by przy wgrywaniu przez użytkownika danych do systemu identyfikować je po dacie, godzinie, minucie i sekundzie ich zapisu na serwerze. Wszystko działało doskonale przez długie lata, a firma (i ilość klientów) stabilnie rosła. Aż traf chciał, że w końcu jednego dnia, o jednej godzinie,

minucie i sekundzie dwóch różnych użytkowników wgrało na serwer swoje dane. System zidentyfikował ich – zgodnie z programem – po dacie godzinie minucie i sekundzie operacji – jako ten sam podmiot, więc danymi klienta „A" nadpisał dane klienta „B", generując tym samym incydent ochrony danych osobowych.

Jaki z tego wniosek?
Przypadki brzegowe mają ten niefortunny zwyczaj, że się zdarzają – podobnie jak zdarzają się inne rzeczy w nieśmiertelnym cytacie z filmu „Forrest Gumpa"[8]. Czy twierdzę że wszystkie należy traktować jako pewne i podejmować nierzadko drogie działania w celu ich uniknięcia? Oczywiście że nie. Ale naprawdę warto mieć świadomość tego, że mogą wystąpić i brać to pod uwagę podczas prac projektowych i szacowania ryzyka. I o ile przeznaczanie wielkich sum na zabezpieczenie się przed każdą sytuacją, która naprawdę może nigdy nie wystąpić może być niecelowe, to na pewno warto zainwestować w opracowanie planów działania na wypadek wystąpienia takich sytuacji. Oczywiście wiadomo, że najlepiej jest zapobiegać, ale we współczesnych realiach gospodarczych niewiele gorszym rozwiązaniem może być solidny i przetestowany plan ciągłości działania oraz reagowania na incydenty. I dobrze udokumentowana realizacja obowiązku stosowania zasady privacy by design, która w razie wystąpienia incydentu pozwoli podjąć jakąś obronę przed Urzędem Ochrony Danych Osobowych.

Przypisy:
[1] https://uodo.gov.pl/pl/138/2387 [dostęp 07.08.2024 r.]
[2] https://uodo.gov.pl/decyzje/DKN.5131.33.2021 [dostęp 07.08.2024 r.]
[3] https://www.rp.pl/rynek-pracy/art40928641-kobiety-polubily-sztuczna-inteligencje-szczegolnie-polki [dostęp 07.08.2024 r.]
[4] https://noyb.eu/pl/noyb-urges-11-dpas-immediately-stop-metas-abuse-personal-data-ai [dostęp 07.08.2024 r.]
[5] https://technologia.dziennik.pl/aktualnosci/artykuly/8738811,chatgpt-openai-hakerzy-wyciek-danych-cyberbezpieczenstwo.html [dostęp 07.08.2024 r.]
[6] https://www.i-programmer.info/news/149-security/8548-reboot-your-dreamliner-every-248-days-to-avoid-integer-overflow.html [dostęp 07.08.2024 r.]
[7] Z angielskiego taki błąd to „*integer overflow*".
[8] https://forrestgump.fandom.com/wiki/Shit_happens [dostęp 07.08.2024 r.]

Daniel Taberski

Radca Prawny, Specjalista ds. ochrony danych osobowych