e-prywatność w platformie do prowadzenia konsultacji społecznych realizowanej w projekcie *W dialogu*

Piotr Andruszkiewicz

17 listopada 2015

Wprowadzając narzędzia informatyczne, które mają służyć mieszkańcom, należy mieć na uwadze, jakże ważną w dobie społeczeństwa informacyjnego, e-prywatność [1]. Jest to wyjątkowo istotne w przypadku wykorzystywania tego typu narzędzi przez organizacje rządowe [2]. Platforma do prowadzenia konsultacji społecznych realizowana w ramach projektu *W dialogu* ma służyć i pomagać urzędom miast w procesie realizowania takich konsultacji. Niezwykle ważną rolę e-prywatności podkreśla Coleman [3] i wymienia trzy sposoby interpretowania e-prywatności:

1. prywatność użytkownika – użytkownik, mieszkaniec w przypadku konsultacji społecznych nie może zostać zidentyfikowany, tzn. nie powinno być możliwośći ustalenia danych personalnych, np. adresu, użytkownika zabierającego głos w konsultacjach społecznych.

2. prywatność danych – dane dostarczane przez użytkowników powinny być agregowane, a indywidualne wartości przekazywane przez użytkowników powinny być anonimizowane.

3. prywatność usługi – usługodawca powinnien zawrzeć jasny i zrozumiały, także niełamiący prawa, kontrakt dotyczący zbierania i późniejszego wykorzystywania danych.

Według Colemana do efektywnego działania narzędzia informatycznego wspomagającego komunikację instytucji rządowych i mieszkańców potrzebne jest wdrożenie co najniej jednej interpretacji e-prywatności [3].

Rozpoczynając dyskusję na temat e-prywatności w odniesieniu do platformy do prowadzenia konsultacji społecznych realizowanej w ramach projektu *W dialogu*, należy zaznaczyć, że każdy użytkownik, również urzędnik, posiadać będzie swoje odrębne i poświadczone konto, do którego będzie musiał się zalogować, aby przygotować debatę bądź brać w niej udział. Uprawnienia danego użytkownika zależne są od pełnionej roli.

Rozpatrując pierwszą interpretację e-prywatności w kontekście platformy do prowadzenia konsultacji społecznych realizowanej w ramach projektu *W dialogu*, należy zwrócić uwagę na proces rejestracji. Użytkownik podczas rejestracji może podać, nie jest to obowiązkowe, dane o sobie, które stanowić będą jego profil, np. data urodzenia, miejsce zamieszkania, i posłużą do wyboru danego mieszkańca do konkretnej konsultacji społecznej zgodnie z jej warunkami, np. kosultacja

społeczna może dotyczyć mieszkańców konkretnej dzielnicy. Dane podawane przez użytkownika podczas rejestracji (profil użytkownika) są odizolowane od danych zawierających treść debat i wyniki ankiet; znajdują się one w innych tabelach bazy danych. Dostęp do danych rejestracyjnych posiadają tylko wybrani urzędnicy. Dostęp ten jest niezbędny, aby mogli oni wybrać do konsultacji mieszkańców zgodnie z założonymi kryteriami i wysłać zaproszenia. Przykładowymi kryteriami może być pełnoletność i zamieszkiwanie w określonej dzielnicy. Mieszkaniec natomiast ma jedynie dostęp do własnych danych rejestracyjnych.

Uczestnicy konsultacji przeprowadzanej przy użyciu platformy nie są anonimowi z punktu widzenia pozostałych uczestników, tzn. identyfikowani są poprzez pseudonim (ang. nick) bądź imię i nazwisko, ewentualnie imię i pierwszą literę nazwiska. Wyboru wariantu identyfikacji uczestników dokonuje urzędnik, biorąc pod uwagę liczność uczestników debaty i ustalając ją w ten sposób, aby nie było możliwe wywnioskowanie danych osobowych konkretnego uczestnika, np. jego adresu.

Podczas prowadzenia debat tekstowych i głosowych, podobnie jak w przypadku debat bezpośrednich, uczestnicy także muszą sami kontrolować informacje o sobie, które dostarczają. Urzędnicy prowadzący konsultacje społeczne przy użyciu platformy nie będą w stanie zapobiec sytuacjom, w których uczestnik świadomie bądź nieświadomie poda swoje dane osobiste do wiadomości pozostałych uczestników. Przypadek ten jest sytuacją analogiczną do zdarzeń, które mogą mieć miejsce podczas prowadzenia debaty bezpośredniej, np. uczestnik debaty może stwierdzić: „A ja mieszkam na ulicy Kubusia Puchatka 23 i tam kolor kamienicy jest przepiękny, bo łososiowy".

Podsumowując powyższe rozważania, platforma realizuje postulaty pierwszej interpretacji e-prywatności, a więc spełnia kryteria efektywności wg Colemana.

W przypadku prywatności danych platforma realizuje ten postulat m.in. poprzez brak odniesienia indywidualnych danych przekazywanych przez użytkownika w ankietach do danych rejestracyjnych. Podobnie brak jest dostępnego dla uczestników debat połączenia wypowiedzi uczestników debat z danymi rejestracyjnymi. Gromadzone wypowiedzi przypisane są jedynie do uczestnika debaty, do którego danych rejestracyjnych pozostali uczestnicy nie mają dostępu.

Detaliczne wyniki ankiet dostępne są tylko urzędnikom pełniącym odpowiednie role. A upubliczniane przez urzędników będą jedynie zagregowane wyniki ankiet, co powoduje, że indywidualne dane dostarczone przez użytkownika nie są upubliczniane.

Aby zwiększyć poziom e-prywatności można także zastosować techniki zapewniania prywatności stosowane w eksploracji danych (ang. data mining). Coleman w [3] w ramach istotnej dla e-prywatności literatury wymienia publikację [4] prezentującą jedną z technik zapewniania prywatności w eksploracji danych. Technika ta należy do grupy technik kryptograficznych zapewniania prywatności w eksploracji danych [5, 6, 7, 8]. Charakteryzują się one występowaniem kilku właścicieli danych, w przypadku konsultacji społecznych byliby to uczestnicy tych konsultacji, którzy chcą zbudować model statystyczny na podstawie posiadanych danych bez przekazywania ich innym stronom w jawny sposób. Taki proces przy każdorazowym wyliczeniu zdefiniowanych wcześniej statystyk wymaga uczestnictwa właścicieli danych (tutaj uczestników konsultacji) w procesie obliczeń. Przy stosowaniu tego typu technik możliwe byłoby wyliczenie zdefiniowanych wcześniej statystyk w jednej rundzie obliczeń. Natomiast, aby wyliczyć dodatkowe statystyki, a taka sytuacja zdarza się niemalże

zawsze w pracy analityka opisującego rezultaty badań, należałoby przeprowadzić proces zbierania danych ponownie, co w wielu przypadkach będzie bardzo utrudnione bądź niemożliwe, biorąc pod uwagę liczbę osób biorących udział w konsultacjach społecznych.

W przypadku gromadzenia danych w scenariuszu ankietowym lepiej sprawdzają się techniki zapewniania prywatności danym scentralizowanym [9, 10, 11]. Proces gromadzenia danych z wykorzystaniem tych technik w uproszczeniu przebiega następująco: użytkownik wprowadza dane do aplikacji, aplikacja zakłóca indywidualne dane podane przez użytkownika i te zakłócone dane wysyła na serwer, gdzie są one centralnie składowane. W ten sposób użytkownik nie przekazuje swoich indywidualnych danych a jedynie zakłócone dane, a administrator danych może wielokrotnie wyliczać na zgromadzonych danych potrzebne statystyki i realizować zadania eksploracji danych. Wadą technik zapewniania prywatności danym scentralizowanym, w szczególności tych opartych na randomizacji, jest spadek jakości uzyskiwanych rezultatów wraz ze wzrostem wprowadzanego poziomu prywatności, zwany kompromisem pomiędzy prywatnością a jakością [12].

Obie wymienione grupy zapewniania prywatności, kryptograficzna i operująca na danych scentralizowanych, stosowane są dla danych numerycznych. Jednakże w przypadku platformy dane tekstowe będą zapisami debat głosowych bądź tekstowych, które z założenia mają być dostępne dla innych uczestników konsultacji.

Analizując prywatność danych w kontekście platformy, można stwierdzić, że realizuje ona podstawowe postulaty e-prywatności w odniesieniu do danych. Natomiast można by było wyposażyć ją w dodatkowe techniki zwiększające poziom prywatności.

W przypadku prywatności usługi strona dostarczająca narzędzia informatycznego powinna zawrzeć z użytkownikami jasny, zgodny z przepisami kontrakt dotyczący zasad gromadzenia i późniejszego wykorzystania danych. Platforma informuje użytkownika o tych zasadach, przestawiając mu podczas rejestracji regulamin, który użytkownik musi zaakceptować. Pozwala więc na zawarcie jasnego kontraktu dotyczącego gromadzenia i późniejszego wykorzystania danych. Zasady zawarte w takim regulaminie pozotają już w gestii urzędów prowadzących konsultacje społeczne z wykorzystaniem platformy. Podobnie jak przygotowywanie rocznych raportów szczegółowo opisujących kroki jakie podjęły urzędy w celu zapewnienia prywatności użytkowników systemu oraz działań podjętych w przypadku wystąpienia ewentualnych naruszeń tej prywatności.

Podsumowując, platforma do prowadzenia konsultacji społecznych realizowana w ramach projektu *W dialogu* realizuje postulaty prywatności użytkownika oraz serwisu. Wypełnia także podstawowe zasady prywatności danych. Umożliwia to więc realizację, konstulacji społecznych z wykorzystaniem wspomnianej platformy przy zachowaniu e-prywatności.

Literatura

[1] Rezgui, A., Ouzzani, M., Bouguettaya, A., Medjahed, B.: Preserving privacy in web services. In: Proceedings of the 4th International Workshop on Web Information and Data Management. WIDM '02, New York, NY, USA, ACM (2002) 56–62

[2] Hiller, J.S., Bélanger, F.: Privacy strategies for electronic government. In: E-government 2001, Rowman & Littkefieks Publishers INC, Citeseer (2001)

[3] Coleman, S.: In dialogue – some observations on a new platform for local government online consultations

[4] Patsakis, C., Laird, P., Clear, M., Bouroche, M., Solanas, A.: Interoperable privacy-aware e-participation within smart cities. Computer **48**(1) (Jan 2015) 52–58

[5] Kapoor, V., Poncelet, P., Trousset, F., Teisseire, M.: Privacy preserving sequential pattern mining in distributed databases. In: CIKM. (2006) 758–767

[6] Jagannathan, G., Pillaipakkamnatt, K., Wright, R.N.: A new privacy-preserving distributed k-clustering algorithm. In: SDM. (2006)

[7] Kantarcioglu, M., Clifton, C.: Privacy-preserving distributed mining of association rules on horizontally partitioned data. In: DMKD. (2002)

[8] Hussein, M., El-Sisi, A., Ismail, N.A.: Fast cryptographic privacy preserving association rules mining on distributed homogenous data base. In: KES (2). (2008) 607–616

[9] Agrawal, R., Srikant, R.: Privacy-preserving data mining. In: SIGMOD Conference. (2000) 439–450

[10] Rizvi, S.J., Haritsa, J.R.: Maintaining data privacy in association rule mining. In: VLDB '02: Proceedings of the 28th international conference on Very Large Data Bases, VLDB Endowment (2002) 682–693

[11] Xia, Y., Yang, Y., Chi, Y., Muntz, R.R.: Mining association rules with non-uniform privacy concerns. Technical Report CSD-TR No. 040015, ftp://ftp.cs.ucla.edu/tech-report/2004-reports/040015.pdf, UCLA (2004)

[12] Agrawal, S., Krishnan, V., Haritsa, J.R.: On addressing efficiency concerns in privacy preserving data mining. CoRR **cs.DB/0310038** (2003)