Gestion des données sur Belenios

17 janvier 2024

Document rédigé par Véronique Cortier, Pierrick Gaudry, et Stéphane Glondu.

Résumé

Ce document a pour objectif de détailler quelles sont les données personnelles stockées sur le serveur Belenios, aux différentes étapes d'une élection, et quel est leur traitement.

Les données stockées ne servent que pour le fonctionnement de l'élection correspondante. Les adresses email ne sont utilisées que pour envoyer le matériel électoral de l'élection aux électeurs. Les seules informations utilisées à des fins statistiques, pour rendre compte de l'utilisation du serveur Belenios, sont décrites au paragraphe 3.3.

Les informations décrites dans ce document correspondent à la version 1.19 du logiciel Belenios, qui implémente la version 1.17 de la spécification.

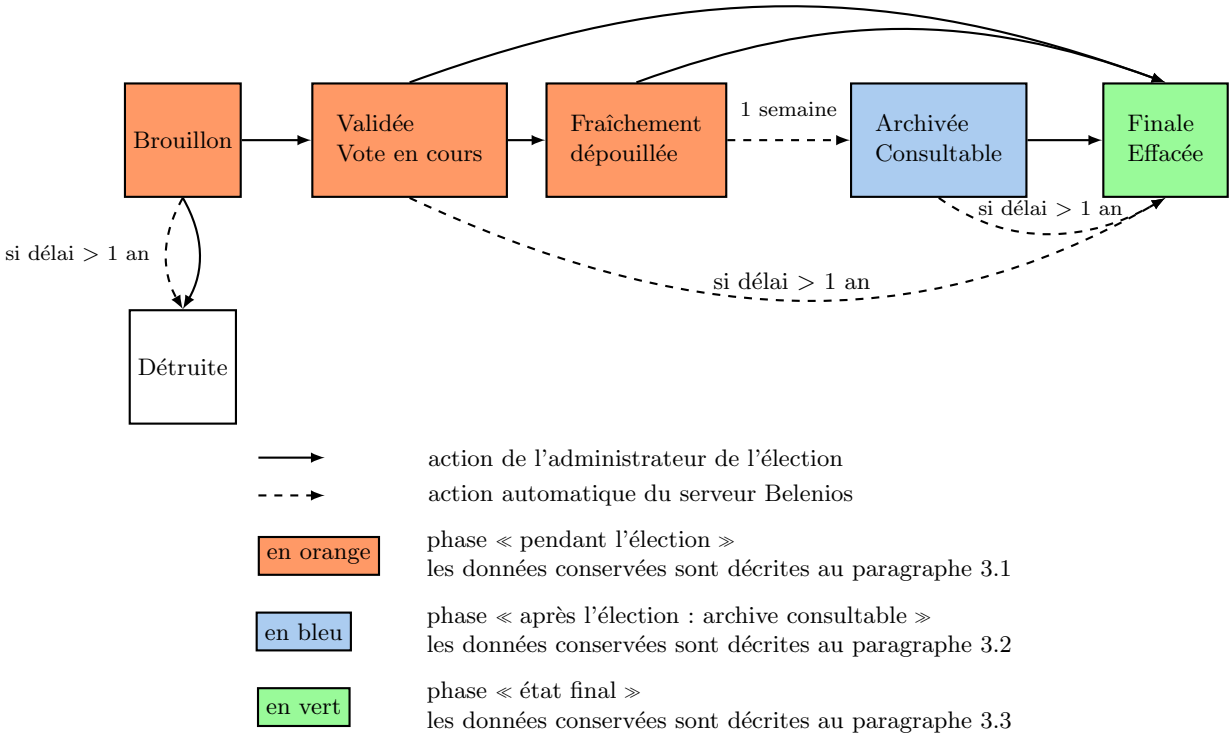
1 Préliminaires

Les élections menées à l'aide de Belenios impliquent plusieurs intervenants, humains ou automatisés. Le point central des interactions est le serveur Belenios, accessible depuis Internet, et qui stocke un grand nombre d'informations. Ce document décrit la gestion de ces données par le serveur, et exclusivement par le serveur. Les données stockées par les tiers (l'administrateur de l'élection, les autorités de déchiffrement, les autorités de jetons de vote) relèvent de leur propre responsabilité. Dans tout ce qui suit, lorsque l'on parle de stockage ou d'effacement de données effectué par Belenios, cela ne concerne que le serveur central.

Les données stockées par le serveur Belenios sont sauvegardées automatiquement à intervalle régulier afin de pouvoir poursuivre le déroulement des élections en cours en cas de corruption des données suite à un problème matériel ou un bug logiciel. Ces sauvegardes sont stockées sous forme chiffrée, sur une autre machine non exposée à Internet, et sont effacées au bout d'une durée d'1 mois.

Nous utilisons des cookies pour enregistrer la préférence de langue (si l'électeur choisit une langue différente de celle par défaut), pour enregistrer le consentement de l'électeur sur l'utilisation des cookies et pour gérer la phase de vote. Il s'agit de relier les différentes requêtes faites par l'électeur pendant la phase de vote. Une fois que l'électeur a fini de voter, toutes les informations identifiantes sont effacées. En particulier, nous n'utilisons pas les cookies à des fins statistiques ni publicitaires.

2 Évolution type pour une élection



L'administrateur de l'élection décide de l'avancement de l'élection : brouillon, validée, fraîchement dépouillée (à chaque fois, il s'agit de l'état pendant l'élection), archivée consultable (après l'élection), finale effacée (état final).

À tout moment, il peut supprimer les données d'une élection en basculant dans l'état final. Si une élection n'a pas été validée (donc l'élection n'a pas pu commencer), toutes les données sont purement et simplement effacées, au bout d'un an maximum, ou avant sur action de l'administrateur. Dès qu'une élection est dépouillée, l'administrateur peut télécharger une archive contenant :

— les données publiques (cf Section 3),
— les données administrateur (cf Section 3).

Cette archive peut être conservée par l'administrateur suivant le besoin de l'élection, tout en supprimant les données sur le serveur (en basculant dans l'état final). Sans action de l'administrateur, le serveur Belenios supprime les données d'une élection (état final) au bout d'un an sans changement d'état. D'autre part, une semaine après l'élection, le serveur bascule l'élection dans l'état « après l'élection ».

3 Détail des données stockées par Belenios

3.1 Pendant l'élection

Avant, pendant et 1 semaine après la fin de l'élection (dépouillement), les données suivantes sont conservées sur le serveur.

données publiques (accessibles aux électeurs)

— titre de l'élection
— questions de l'élection
— résultat de l'élection

— clés publiques des autorités
— partie publique des jetons de vote
— bulletins chiffrés
— preuve de bon déchiffrement

données administrateur (accessibles à l'administrateur de l'élection)
— liste électorale (email et login)
— liste d'émargement, horodatée
— dans le mode où le serveur génère les codes de vote (pas d'autorité externe de générateur de code de vote) et avant le début de l'élection (état « Brouillon »), l'administrateur peut copier la partie privée des codes de vote des électeurs. Cela est utile en cas de perte du code de vote par un électeur. Cela ne dégrade pas la sécurité si le serveur de vote est de confiance. Et s'il ne l'est pas, une autorité externe de code de vote doit être utilisée.

données serveur (accessibles aux administrateurs du serveur Belenios)
— mots de passe des électeurs (salés et hachés), lorsque le mode mot de passe est choisi. NB : par défaut, c'est l'authentification par mot de passe éphémère qui est utilisée, sans stockage de mot de passe sur le serveur.
— logs généraux de connection au serveur
— fichier security.log (contient les tentatives échouées et suspectes de revote)
— date de validation et de dépouillement
— clé de déchiffrement : le serveur est toujours l'une des autorités de déchiffrement.
— base de données interne : enregistre le lien entre les électeurs et leur code (publique) de vote.
— dans le mode « chiffrement à seuil » (il suffit de k clés parmi n pour déchiffrer l'élection) : les clés de déchiffrement des autorités (trustees) sont stockées *chiffrées* par la clé publique de l'autorité correspondante.
— dans le mode où le serveur génère les codes de vote (pas d'autorité externe de générateur de code de vote), la partie privée des codes de vote est stockée sur le serveur avant la mise en place de l'élection (état « Brouillon »).

Les administrateurs du serveur ont également accès aux données publiques et aux données administrateur.

Note : les logs généraux de connection au serveur et le fichier security.log sont gérés en dehors de la base de données de Belenios. De ce fait, leur gestion est différente : indépendamment des choix faits par l'administrateur de l'élection, ces données sont conservées 2 semaines sur le serveur puis effacées.

3.2 Après l'élection : archive consultable

Après le dépouillement (et pendant 1 an), l'élection est archivée. Il est possible pour les électeurs de consulter le résultat et de vérifier sa cohérence vis-à-vis des bulletins chiffrés de l'élection. Il est également possible pour l'administrateur de l'élection de consulter la liste d'émargement. Les données conservées sur le serveur sont toutes celles mentionnées précédemment (paragraphe 3.1), **sauf** :
— la clé de déchiffrement (si le serveur possédait une des clés ou la clé de déchiffrement)
— le lien entre les électeurs et leur code (publique) de vote, stocké dans la base de données interne.
— dans le mode « chiffrement à seuil » (il suffit de k clés parmi n pour déchiffrer l'élection) : les clés de déchiffrement des autorités (trustees) étaient stockées *chiffrées* par la clé publique de l'autorité correspondante. Ces chiffrés sont désormais effacés.

3.3 État final

1 an après le dépouillement (ou à défaut, 1 an après la création de l'élection), les données de l'élection sont effacées. Seules quelques informations sont conservées à des fins statistiques, dans le

but de rendre compte de l'usage du serveur de Belenios. L'administrateur de l'élection peut choisir d'effacer l'élection (passage dans l'état final) avant ce délai d'un an.

Les données conservées à des fins statistiques sont les suivantes.

— titre de l'élection
— la forme des questions (type d'élection, nombre de réponses et nombre de choix), toutes les chaines de caractères sont remplacées par la chaine vide. Les questions posées et les choix possibles sont donc effacés
— nombre d'électeurs et nombre de bulletins enregistrés
— identifiant de l'organisateur (suivant le mode d'authentification choisi, il peut s'agir de son adresse mail ou d'un login)
— date du dépouillement (ou, à défaut, validation de l'élection)
— choix du mode d'authentification et, le cas échéant, l'adresse du serveur CAS utilisé
— choix du mode de gestion des jetons de vote
— concernant les autorités de dépouillement : leur nombre et la présence ou non du serveur parmi elles
— l'élection a-t-elle été dépouillée ?

4 Sensibilité des données stockées

Voici notre analyse (subjective) de la sensibilité des données stockées sur la plateforme de vote Belenios avec une brève analyse des risques associés.

Extrêmement sensible

— La clé de déchiffrement lorsqu'elle est stockée sur le serveur et qu'aucune autre autorité de déchiffrement n'a été choisie (mode sans autorités de déchiffrement). Avec les données conservées sur le serveur, il est alors possible de lire qui a voté quoi. Ce mode est permis à des fins de test ou pour des élections absolument sans enjeu (choix de la pizza pour ce soir) et ne devrait pas être utilisé pour des élections sérieuses.

Scénarios d'attaque : le serveur est piraté (physiquement ou à distance) ou les administrateurs du serveur sont corrompus.

Très sensible

— Base de données interne, permet de faire le lien entre un électeur et son bulletin chiffré.

Le risque est la perte de confidentialité du vote si de plus la sécurité du chiffrement n'est plus assurée.

Scénarios d'attaque : intrusion sur le serveur ET les clés de déchiffrement fuitent (risque à évaluer en fonction des trustees) ou la crypto est cassée (ordinateur quantique ?).

— La clé de déchiffrement lorsqu'elle est stockée sur le serveur et que d'autres autorités de déchiffrement ont été choisies (rappel : lorsque le serveur est la seule autorité de déchiffrement, on est dans le cas extrêmement sensible, cf supra).

Scénarios d'attaque : intrusion sur le serveur ET les autres clés de déchiffrement fuitent (risque à évaluer en fonction des trustees) ou la crypto est cassée (ordinateur quantique ?).

— Les parties privées des codes de vote des électeurs stockées sur le serveur avant le début de l'élection, lorsqu'il n'y a pas d'autorité externe de codes de vote. Le risque est l'ajout de bulletins (bourrage d'urne).

Scenarios d'attaque : intrusion sur le serveur avant l'élection pour récupérer les parties privées des codes de vote ET intrusion sur le serveur pendant l'élection pour ajouter des bulletins. Cette attaque peut être détectée si l'attaquant ne met pas correctement à jour la liste d'émargement ou, peut-être, par un examen attentif de la liste d'émargement (e.g. des personnes qui n'ont pas voté apparaissent sur la liste d'émargement).

Sensible

— Liste d'émargement : permet de faire le lien entre électeurs et bulletins chiffrés SI l'attaquant a surveillé les temps d'arrivée des bulletins sur l'urne. Permet aussi de savoir qui a participé à quelle élection et à quelle heure.
— Logs généraux de connection au serveur (ip, user-agent, date, url) : assez similaire à la liste d'émargement.

  Le risque est de nouveau la perte de confidentialité du vote si de plus la sécurité du chiffrement n'est plus assurée.

  Scénarios d'attaque : fuite de la liste d'émargement (ou des logs serveurs) ET monitoring actif pendant l'élection ET les clés de déchiffrement fuitent (risque à évaluer en fonction des trustees) ou la crypto est cassée (ordinateur quantique ?).
— Les clés de déchiffrement des autorités (trustees), stockées *chiffrées* par la clé publique de l'autorité correspondante (dans le mode « chiffrement à seuil » uniquement) : permet de déchiffrer tous les bulletins SI l'attaquant réussit par ailleurs à obtenir les clés privées d'un nombre suffisant d'autorités.

  Scénarios d'attaque : fuite de la liste d'émargement (ou des logs serveurs) ET monitoring actif pendant l'élection ET fuite de suffisamment de clés privées des autorités de déchiffrement (chargées de stocker leur clé par leurs propres moyens) ou la crypto est cassée (ordinateur quantique ?).

Assez sensible

— Liste électorale (= liste d'emails).

  Le risque est une utilisation de cette liste d'emails pour un autre contexte (spam, phishing ciblé).
— Base de mots de passe hachés (contient également la liste des emails), lorsque que l'authentification par mot de passe est utilisée, ce qui n'est pas l'option par défaut.

  Les mots de passe sont générés par le serveur et non pas par l'utilisateur, donc le risque associé est minimal.