Consentimento informado

Responsáveis pelo projeto
Prof. André Zúquete (andre.zuquete)
Prof. Tomás Oliveira e Silva (tos@ua.pt)

1 Introdução

A UA dispõe de uma infraestrutura de rede Wi-Fi que permite registar todas as associações de terminais móveis com os pontos de acesso à rede do campus. Esta capacidade permite saber de forma bastante aproximada o local onde está o terminal de uma pessoa, o que, assumindo que o mesmo está acompanhado pelo dono, permite saber uma localização aproximada dessa pessoa (edifício, talvez sala).

Esta capacidade pode ser alavancada para saber, com alguma precisão, se um(a) aluno/a esteve presente ou não numa determinada aula. Embora esta informação seja de pouca valia para as aulas onde são registadas as presenças de alunos, ela é relevante para as aulas onde tal registo não é realizado. Assim, esta capacidade pode ser usada para fornecer um mecanismo de gestão relevante para conhecer a taxa de ocupação das salas de aula, em especial nas em que não se façam registos presenças.

A informação de localização dos terminais Wi-Fi pessoais é informação pessoal, sensível, e o seu processamento implica especiais cuidados que garantam a privacidade dos titulares dessa informação pessoal. Neste sentido, o que se pretende, como produto final, não é saber se um determinado aluno esteve numa aula específica, mas saber o seguinte:

- Quanto alunos de uma turma estiveram numa aula dessa turma;
- Qual a distribuição de percentagens de assiduidade dos alunos numa aula.

Esta informação não permite qualquer identificação dos alunos, pelo que não afeta a sua privacidade. Mais, esta informação será usada apenas para consumo interno da UA, e não será nunca exposta publicamente.

2 Aproximação tecnológica para a proteção da privacidade de dados sensíveis

Para implementar o sistema será usada uma tecnologia de computação especial, denominada Intel SGX. Esta permite criar enclaves, que são universos de computação confidenciais. Os enclaves SGX não permitem que as computações neles realizadas possam ser observadas seja por quem for, sendo esta proteção garantida pelo processador.

Os enclaves SGX comunicam de forma limitada com o exterior, usando para o efeito duas APIs:

- A API que permite receber pedidos ou dados do exterior (i-calls);
- A API que permite fazer pedidos ou enviar dados para o exterior (o-calls).

Assim, um problema que se coloca é se um enclave não pode divulgar dados privados de forma inapropriada através de resultados de i-calls ou como parâmetros das o-calls.

Este problema é resolvido através de um processo designado por atestação. A atestação permite verificar se o software do enclave SGX corresponde a uma determinada versão do software. Sempre que o software mudar, a atestação deteta essa mudança. Combinando a atestação com um processo de certificação, onde o código do enclave SGX pode ser analisado por entidade independente para verificar que o mesmo não liberta qualquer informação pessoal de qualquer forma, consegue-se garantir a manipulação confidencial dessa informação pessoal.

Complementarmente, um enclave SGX pode produzir internamente pares de chaves criptográficas assimétricas (RSA). Se o enclave proteger de forma conveniente a chave privada destes pares, todos os dados cifrados com as correspondentes chaves públicas no exterior do mesmo ficam apenas acessíveis, em termos de compreensão, para o enclave. Ou seja, com estes pares de chaves consegue-se garantir que dados sensíveis possam permanecer cifrados no exterior do enclave e possam ser processados exclusivamente pelo enclave.

Concluindo, combinando o uso de enclaves SGX bem desenhados e certificados é possível criar volumes de dados com elementos sensíveis protegidos por uma cifra que só podem ser processados de forma confidencial pelos enclaves.

3 Fontes de dados a processar e anonimização dos dados pessoais

No desenho, desenvolvimento, teste e exploração do sistema são usadas duas fontes de dados reais complementares com informação pessoal:

1. O horários das turmas a analisar e os seus alunos (só interessa o Utilizador Universal); Estes serão obtidos numa base semanal para atender a mudanças de turma ao longo do semestre.
2. Os registos diários de acessos à rede Wi-Fi.

3.1 Fase de exploração

Na fase os vários registos serão imediatamente cifrados com a chave pública do enclave SGX assim que são extraídos de outros sistemas da UA (PACO e Cisco Catalyst, respetivamente).

Nesta fase apenas estarão envolvidas pessoas dos STIC com acesso privilegiado a esses dois sistemas.

3.2 Fase de desenho, desenvolvimento e testes

Esta fase, que é a que tem relevância para este consentimento informado, envolverá uma população de alunos minimizada de forma a não levantar problemas de privacidade perante o investigador (aluno finalista do mestrado em Cibersegurança da UA) que conceberá e testará o sistema. Assim, os dados pessoais envolvidos serão minimizados e anonimizados de forma irreversível pelas pessoas dos STIC que realizarem a sua extração seguindo os procedimentos que abaixo se descrevem.

4 Minimização da informação com dados sensíveis

A minimização consistirá apenas em considerar um conjunto de turmas-alvo sem alunos comuns e em descartar os registos de acesso à rede garantidamente irrelevantes. Mais ainda, a minimização passará pelo uso do Utilizador Universal dos alunos dessas turmas que **explicitamente** concordarem explicitamente em participar neste projeto de desenvolvimento.

Como nesta fase se vão usar apenas turmas que tenham aulas em salas do DETI, vamos usar apenas as turmas de **Arquitetura de Computadores (43673)** e as turmas de **Arquitetura de Computadores I (41948)**. Mais nenhuma turma será envolvida, o que impede qualquer desanonimização dos dados através do cruzamento da mesma pessoa em diferentes turmas. Da constituição destas turmas serão **excluídos** todos alunos que não concordarem em participar neste projeto de desenvolvimento.

Relativamente aos alunos que concordarem com o estudo, só vão ser considerados os seus registos de acesso à rede no dia em que a sua turma funciona, e não em qualquer outro dia.

5 Anonimização dos dados pessoais

A anonimização tanto das turmas como dos registos de acesso à rede consistirá na **transformação unidirecional irreversível** do Utilizador Universal de cada aluno envolvido no estudo.

Esta transformação é feita com recurso a funções criptográficas apropriadas (funções de dispersão unidirecional criptográfica, também conhecidas como funções de síntese, como a SHA-256) e a um segredo (senha) conhecido apenas das pessoas dos STIC que realizaram a extração dos dados originais. Para tornar o processo mais robusto, o sistema de cálculo dos identificadores anónimos usará uma função arbitrariamente lenta, como a PBKDF2 ou a Argon2, de forma a dar mais garantias de impossibilidade de reversão da anonimização por via de processos de tentativa-e-erro.

Concluindo, em caso algum o aluno que irá desenhar, desenvolver e testar o sistema terá acesso ao Utilizador Universal dos alunos envolvidos no projeto de desenvolvimento, nem a perfis de acesso às redes fora dos dias de interesse para cada aluno, nem as turmas sobre as quais testará o seu sistema lhe permitem fazer qualquer ataque de inferência sobre os dados anonimizados das turmas e dos acessos à rede Wi-Fi.

6 Validação do sistema

No desenvolvimento do sistema há duas fases de validação:

1. Aquela em que o aluno testará se os resultados produzidos pelo enclave SGX estão de acordo com os dados processados;
2. Aquela em que o aluno verificará se os resultados produzidos estão de acordo com a realidade registada pelos docentes das turmas envolvidas.

Na primeira fase o aluno trabalhará exclusivamente com os dados anonimizados da forma acima descrita. A anonimização é um requisito fundamental, portanto.

A segunda fase consistirá na obtenção, por pessoas responsáveis dos STIC, de um ficheiro por cada aula de cada turma envolvida com pares (Utilizador Universal, falta). Estes pares apenas envolverão os alunos que participam no estudo, e o seu Utilizador Universal será anonimizado da forma antes descrita. Em caso algum serão envolvidos os docentes dessas turmas. Estes dados serão facultados ao aluno só após a sua minimização e anonimização.

7 Destruição dos dados anonimizados

Os dados anonimizados serão produzidos durante o decurso do primeiro semestre e permanecerão guardados apenas durante a fase de desenho, desenvolvimento e teste do sistema. Após esse período, serão destruídos sem recuperação possível, tanto pelas pessoas dos STIC responsáveis pela sua criação, como pelo aluno que os irá usar para desenvolver o sistema.

8 Implicações do consentimento ou não-consentimento de cada discente

Nenhum discente será beneficiado por consentir em participar neste projeto de desenvolvimento. Se consentir, mamtém o direito a retirar o seu consentimento a qualquer momento (Art. 7, n. 3 do RGPD). Finalmemnte, se consentir, tem o direito de apresentar uma reclamação à autoridade de controlo (CNPD, Art. 77 do RGPD) caso considere justificável.

Nenhum discente será prejudicado por não consentir em participar neste projeto de desenvolvimento, expressando dessa forma o seu direito de oposição (Art. 21 do RGPD).

9 Manifestação de consentimento

Eu, (nome) _____

discente da UA com o Número Mecanográfico _____

e com o Utilizador Universal (e-mail) _____

tomei conhecimento deste projeto de desenvolvimento de um sistema de obtenção de dados da ocupação de aulas e da distribuição de frequências de assiduidade por turma, que será desenvolvido e testado com os meus dados minimizados e anonimizados. Mais tomei conhecimento de como e por quem é que os meus dados serão extraídos, minimizados, anonimizados, processados, usados para validação e eliminados.

Consequentemente, indico que concordo com a minha inclusão no mesmo.

data: _____

Assinatura: _____

data: _____