Design of a Privacy-preserving Document Submission and Grading System

Benjamin Greschbach, Guillermo Rodríguez-Cano,
Tomas Ericsson, and Sonja Buchegger*

KTH Royal Institute of Technology, Stockholm, Sweden. {bgre,gurc,te,buc}@kth.se

Abstract Document submission and grading systems are commonly used in educational institutions. They facilitate the hand-in of assignments by students, the subsequent grading by the course teachers and the management of the submitted documents and corresponding grades. But they might also undermine the privacy of students, especially when documents and related data are stored long term with the risk of leaking to malicious parties in the future. We propose a protocol for a privacy-preserving, anonymous document submission and grading system based on blind signatures. Our solution guarantees the unlinkability of a document with the authoring student even after her grade has been reported, while the student can prove that she received the grade assigned to the document she submitted. We implemented a prototype of the proposed protocol to show its feasibility and evaluate its privacy and security properties.

1 Introduction

The pervasive collection of massive amounts of personal data is an increasing threat to user privacy. Often, more information than necessary for the intended purpose is collected and used for profiling or targeted advertisement. User choice is often limited to either not using a given system or service, or to accepting the loss of privacy that comes along with using the system. Designing systems that collect or process personal user data should therefore have privacy in mind from the beginning and employ best practices such as data minimization.

The focus of this work is the context of an educational institution, e. g., a university, where students take courses, work on assignments for these courses and teachers grade these assignments. In this context, discriminatory grading may be an issue, i. e., grading that is not solely based on the student's achievements but also on the teacher's preconception about individual students or stereotypes about certain groups of students. One approach to avoid this is to use blind grading, where the student's identity is not known to the teacher while grading the assignment. Only after the grade has been determined, the link between assignment and student identity is recovered, so that the grade can be assigned to the student. In some settings, one might even want to have what we refer to

* This research has been funded by the Swedish Foundation for Strategic Research grant SSF FFL09-0086 and the Swedish Research Council grant VR 2009-3793

as *forward unlinkability*, i. e., the teacher not being able to link the student to the assignment even after the grades have been reported. For example if a course consists of two different assignments, the work done on the two assignments is linkable if students are likely to choose similar topics for both parts. In that case, without forward unlinkability, the teacher would know the student's identity during the grading of the second assignment. Another motivation for wanting forward unlinkability is the general aim of data minimization, which among other things protects against unintended leakages of personal data in the future. At the same time, the handling and grading of assignments has to guarantee that a student receives a certain grade if and only if she submitted work that was graded accordingly by the teacher. So while the student identity and the submitted document have to remain unlinkable, we want a *provable linkability* of the student's identity and the received grade.

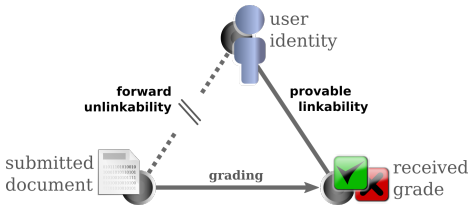

Figure 1. Overview of system entities, their relations and desired properties.

Using the cryptographic technique of blind signatures [1], we propose a protocol for this use case: a privacy-preserving document submission and grading system that allows students to submit documents anonymously without compromising the correctness of the grade assignment process.

After presenting related work in Section 2, we formulate a system model and desired properties for the system in Section 3. In Section 4 we suggest a protocol design that meets these requirements, and evaluate the proposed protocol, discussing its privacy and security properties in Section 5. Furthermore, we show the practical feasibility of the proposed solution by having implemented a proof-of-concept prototype of the protocol, briefly described in the same section.

2 Related Work

Blind signatures schemes are widely used to enhance the privacy of protocols by providing unlinkability. Examples of their use include identity management in federated login systems, e. g., PseudoID [2], a project to protect the login data from the identity providers by means of blind digital signatures, or electronic payment systems, e. g., Taler [6], a digital currency approach close to Bitcoin with the additional benefit of governmental tax traceability without losing anonymity as blind signatures provide unlinkability to the transactions

between customers and merchants but not between government and merchants. Secure voting schemes are another application area of blind signatures, e. g., CryptoBallot [7], a cryptographically secure online voting system where ballots cannot be traced back to the voter as they are blinded but their counting and the voter identities are publicly auditable. Even though these schemes have similarities with our problem, they would be unnecessarily complex to adapt for the use case of document submission and grading. Attribute-based anonymous credentials can be used for similar purposes, such as in an anonymous course evaluation system for universities [9]. The use-case of this project differs, however, from our problem statement, having a focus on smart-card based anonymous course attendance verification, introducing complexity not needed in our scenario. Whistleblower platforms, e. g., SecureDrop [5] or GlobaLeaks [8], allow a sender to submit documents anonymously to a receiver, such as a media organization. These systems employ anonymous and confidential communication and meta-data footprint minimization to increase the anonymity of the sender. While maximizing the sender anonymity, they lack, however, the provable linkability of feedback (grades in our scenario) to identifiers, that is required in our use case.

3 Anonymous Document Submission System

We aim to design a document submission and grading system, where each student can submit a document to the system before a public deadline. After the deadline passed, the teacher grades all submitted documents with either pass or fail. When all documents are graded, each student receives the grade that the teacher assigned to the document that was submitted by the student.

We assume that students can store credentials they receive in a secure way, do not pass them on to others and that they can communicate with the system in a mutually authenticated and confidential way (e. g., via a TLS secured web login), and at other times in an anonymous and confidential way (e. g., by using TLS over Tor[3]). We assume that they are careful to include no identifying information in the documents and that authorship attribution by stylometry is not feasible for the adversary. When discussing the security and reliability of the system from the teacher's perspective, we assume that the server cannot be compromised and that the teacher can handle secret keys in a secure way. Furthermore, protection against ghost writing is out of scope of this work, so we assume that students will not ask someone else to write their documents, which reflects a general limitation for home assignments that are allowed to be worked on outside a teacher-controlled environment. To achieve anonymity and correctness, we want the system to have the following two properties:

student–document forward unlinkability

A document cannot be linked to a student by anyone else than the student who submitted the document, and the unlinkability remains even after grades have been assigned to students.

student–grade provable linkability

If and only if a document was graded with a certain grade, the student who submitted the document can prove that she received this grade.

We want our system to both protect the student's privacy and to protect the teacher from dishonest students. Therefore we consider two different adversaries. The first adversary tries to break the student's anonymity and is capable of compromising any involved party except for the student herself. In particularly it can control the teacher, the server and any other student. Furthermore, we assume this adversary to be capable to passively intercept all network traffic and actively inject messages. The second adversary tries to break the correctness of the grade assignment and is used when discussing the security and reliability of the system from the teacher's perspective. This adversary is assumed to be able to compromise one student, to passively observe all network traffic and to actively inject messages.

4 Protocol Design

We implement the protocol that has the desired properties using a blind signature scheme as described in [1], that provides the functions $blind, unblind, sign$ and $verify$, with the property that blinding perfectly hides the data, but signatures on blinded data can still be verified after unblinding (informally: $unblind(sign(blind(x))) = sign(x)$). Figure 2 shows the sequence of steps in our proposed protocol. First, the system server provides each registered student in the course with a unique, random, one-time identifier rID and stores the relation of student identifiers to rIDs for later use. Next, the student blinds the rID for both the pass verification key e_{pass} and the fail verification key e_{fail}, using two private, random blinding factors b_{pass} and b_{fail}, and sends the resulting $bID_{pass} = blind(rID, b_{pass}, e_{pass})$ and $bID_{fail} = blind(rID, b_{fail}, e_{fail})$ together with D, the document to submit, to the server over an anonymous, encrypted channel. At this point, the server does not learn who submitted the document because the blinding hides the rID, using the anonymous channel obfuscates the network address origin and the document D is assumed to not contain any identifying information about the student. After the deadline has passed, the teacher grades all submitted documents. If a document is graded as passed, the blinded identifier bID_{pass} that was submitted together with the document is signed with the private pass signing key d_{pass} of the teacher. If the grade is fail, bID_{fail} will be signed with the teacher's private fail signing key d_{fail}. When all documents are graded, the server publishes a list of all signed blinded identifiers. The student fetches the list, picks the signed blinded identifier that belongs to her and unblinds it. The student will try both public verification keys e_{pass} and e_{fail} to check which grade she received. By this, the student obtains a signed identifier $sID = sign(rID)$ that proves that she received the grade corresponding to the signing-key. Finally, she sends sID to the server, the server

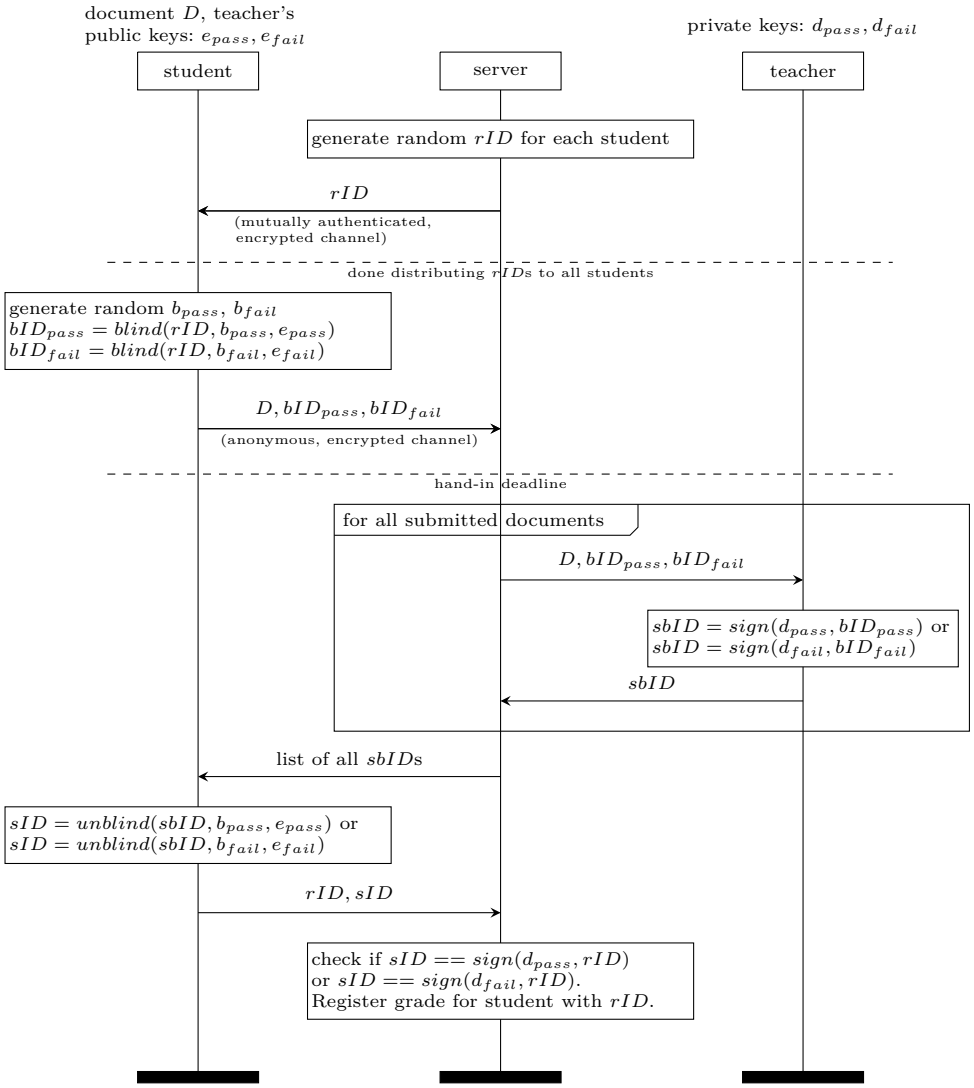

Figure 2. Protocol realizing an anonymous document submission and grading system.

checks the signature, looks up the student identifier that belongs to rID, and registers the corresponding grade for the student, not learning which document the student submitted.

5 Discussion and Evaluation

We have implemented a proof-of-concept prototype (see `http://www.ter.se/dss/`) which is a collection of C programs for the various operations performed by the

student, server and teacher, realizing the described protocol. The prototype uses the free cryptographic library `Libgcrypt` [4]. In the following, we do an informal security and privacy evaluation, discussing why the previously defined properties hold and various attacks will not succeed. We do not cover implementation-based attacks though, such as cross-site-scripting attacks on a web-interface.

Student–document forward unlinkability The student identity is directly linked to the random identifier rID. But when submitting the document D, the rID is perfectly hidden by the random blinding factors known only to the student. We assume that there is no identifying information contained in the document, so after submission the server cannot link student identifiers to documents. To achieve forward unlinkability, this has to hold even after the grades have been assigned to students. During grading, documents are linked to grades, which is a binary domain in our case. The grade information is attached to the blinded rID in form of a signature with one out of two keys, that can be transformed in a verifiable signature on the unblinded rID only by the student. The properties of the blind signature scheme provide the unlinkability between this unblinded signature and the blinded data that was submitted together with the document. At the same time, the signature is provided to the server together with the rID when the students claim their grades, so they provide an unambiguous mapping from every rID to a grade. As a consequence, we do not get perfect unlinkability of student identifiers and documents, but only *k-anonymity*, where k is the number of students who received the same grade. This is a general limitation of every system where the grading party is not trusted and the assignment of grades to student identifiers is verifiable. A worst-case example is a situation where only one student received the grade "fail", so the teacher can infer that the only document she graded with "fail" must belong to this student. Another limitation is the fact, that the anonymity for all students with a certain grade is reduced whenever one student with the same grade gives up their anonymity voluntarily or becomes compromised by the adversary.

Student–grade provable linkability The provable linkability of student identifiers to grades has two directions: (a) soundness: if a student can prove that she received a certain grade, then she must have submitted a document that was graded accordingly, and (b) completeness: if a document submitted by a student was graded with a certain grade, then the student can prove that she received that grade. We show (a) by contrapositive, so we assume that the student did not submit a document that was graded with the grade that the student claims. Now we see that the student cannot prove that she received that grade: she cannot present a valid signature on the rID assigned to her, made with the private key corresponding to the grade, because the private signing key was used by the teacher only to sign other blinded rIDs that were submitted together with documents that were graded accordingly. For (b) we see directly that if a submitted document was graded with a certain grade, the teacher put a signature on the blinded rID submitted together with the document and publishes that later on. So the student can derive a valid signature on her rID by unblinding

the published information and therefore can prove to anyone knowing the public verification key, that she received that grade.

Timing and Correlation Attacks To avoid timing attacks, it is important that certain events in the protocol do not happen before others. For example, the hand-in of documents must not start before all students received their rIDs (denoted by the first dashed line in the message sequence chart in Figure 2), otherwise the anonymity set for the submitting students is immediately reduced to those who already received their rID. For a similar reason it is important that the server publishes the result list with the signed, blinded rIDs after the hand-in deadline and as a complete list. The latter is important, because if students for example would request their individual entries without downloading the complete list, the server could correlate these requests for specific entries (that the server can link to documents) with requests for registering a grade (which contain the identifier rID), that might happen shortly after each other. End-to-end traffic correlation attacks are also relevant for the concrete implementation of the anonymous channel. Tor, for example, does not protect against an adversary that can observe both traffic going into the Tor network and traffic coming out of it [3], so the students should for example be advised not to use the university network when submitting their documents, because it is likely that the same party operates both the university network and the system server and therefore could observe both ends of the students' connections.

Impersonation and Replay Attacks To avoid impersonation and replay attacks, it is important not to use a public or permanent student identifier such as the students' e-mail addresses. Otherwise, an attacker could impersonate a student, e. g., to damage her reputation by submitting a low-quality document in her name. Therefore, we use the unique, random, one-time identifier rID and distribute it over a mutually authenticated and encrypted channel to the student. This makes impersonation without the cooperation of the student impossible because the attacker does not know which rID was assigned to a student. It also prevents replay attacks, as the rID binds the messages both to the student and to the current course, because even if the same protocol is used for several courses, the same student will receive a new rID in each new protocol run.

Attacks on Combined Cryptographic Primitives Attacks on the used cryptographic primitives, such as public key cryptography, hash functions and blind signatures, are out of scope of this work, so we assume them to be secure. However, we have to be careful to use these tools in a secure way, especially when combining them with each other. It is, for example, important to use two different blinding factors b_{pass} and b_{false} when blinding the rID. Otherwise, if for ease of implementation one would use only one common blinding factor b for both blindings, the server would be able to mount the following de-anonymization attack: The server generates specially prepared public keys $e_{pass} = \langle N, e\rangle$ and $e_{fail} = \langle N, e'\rangle$ with the property that both share the same modulus N and the public exponents have a difference of one: $e - e' = 1 \pmod N$. If the student blinds her rID

for these keys using a common blinding factor b, she will submit the following two values to the server: $bID_{pass} = blind(rID, b, e_{pass}) = rID \cdot b^{e} \pmod N$, $bID_{fail} = blind(rID, b, e_{fail}) = rID \cdot b^{e'} \pmod N$. Now, the server can simply divide the two values to obtain the blinding factor b: $bID_{pass}/bID_{fail} = (b^{e} \cdot rID)/(b^{e'} \cdot rID) = b^{e-e'} = b \pmod N$. Having learned b, the server can unblind the bIDs and obtain rID, thus having de-anonymized the student.

6 Conclusions and Limitations

We have described a practical application for blind signatures schemes in the context of a document submission and grading system to improve the privacy of students without undermining the correctness of the grading process. We found that it is feasible to implement such a system, qualified only by the limitations derived from the scenario, e. g., that the provided k-anonymity depends on the number of other students who received the same grade, that students can choose not to reveal their grade, and that documents cannot be linked to students even where this might be desired for pedagogical reasons or penalty measures for plagiarism that go beyond grading the work with fail.

The basic protocol described here can be extended with more functionality such as having several teachers do the grading, using more fine grained grading scales (with the limitation that this decreases the anonymity sets), issuing submission acknowledgements or including individual feedback without breaking the anonymity properties.

References

1. Chaum, D.: Blind Signatures for Untraceable Payments. In: Chaum, D., Rivest, R.L., Sherman, A.T. (eds.) Advances in Cryptology, pp. 199–203. Springer US (1982)
2. Dey, A., Weis, S.: PseudoID: Enhancing privacy in federated login. In: Hot Topics in Privacy Enhancing Technologies. pp. 95–107 (2010), http://www.pseudoid.net
3. Dingledine, R., Mathewson, N., Syverson, P.: Tor: The Second-generation Onion Router. In: SSYM proceedings Vol.13. p. 21. USENIX Association, USA (2004)
4. Free Software Foundation (FSF): Libgcrypt. Online (2007), http://www.gnu.org/software/libgcrypt/
5. Freedom of the Press Foundation (FPF): SecureDrop. Online (2013), https://securedrop.org/
6. French Institute for Research in Computer Science and Automation (INRIA): Taler. Online (2015), http://www.taler.net
7. Hayes, P.D.: CryptoBallot. Online (2013), https://cryptoballot.com/
8. Hermes Center for Transparency and Digital Human Rights: GlobaLeaks. Online (2011), https://www.globaleaks.org/
9. Stamatiou, Y., et al.: Course evaluation in higher education: the patras pilot of abc4trust. In: Rannenberg, K., Camenisch, J., Sabouri, A. (eds.) Attribute-based Credentials for Trust, pp. 197–239. Springer International Publishing (2015)