Privacy-Friendly Incentives and Their Application to Wikipedia

Jan Camenisch, Thomas Gross, Peter Hladky, Christian Hoertnagl

Privacy-Friendly Incentives and Their Application to Wikipedia*

Jan Camenisch[1], Thomas Groß[1], Peter Hladky**[,2], and Christian Hoertnagl[1]

[1] IBM Research - Zurich, Rüschlikon, Switzerland
[2] Department of Computer Science, ETH Zurich, Switzerland

Abstract. Double-blind peer review is a powerful method to achieve high quality and thus trustworthiness of user-contributed content. Facilitating such reviews requires incentives as well as privacy protection for the reviewers. In this paper, we present the concept of privacy-friendly incentives and discuss the required properties. We then propose a concrete cryptographic realization based on ideas from anonymous e-cash and credential systems. Finally, we report on our software's integration into the MediaWiki software.

Keywords: anonymous credentials, pseudonyms, e-cash, privacy, anonymity

1 Introduction

We, as users, all rely increasingly on information on the Internet, ranging from stock quotes and financial news to medical information. Also, businesses and organizations (including governments) rely on information on the Internet to make their decisions—including, for instance, court cases and financial investments. It is therefore crucial that this information can be trusted to be correct.

Information provided by organizations is typically considered trustworthy because organizations are trusted to have quality assurance processes in place. Moreover, they can be held liable for publishing incorrect information. An increasing part of the Internet's content is *user-contributed*. Here, assessing the trustworthiness of the information is much more difficult, because the contributing users are typically barely known and can easily be impersonated. Also, as they can hardly be held liable, users sometimes contribute wrong information, on purpose. Such cases range from discrediting other users to manipulating votes or markets, see [25, 26] for examples.

Sites such as Wikipedia try to address this problem by establishing user's reputation. This is normally done by registration and *identification of the users*, sacrificing users' privacy for the quality of their contributions. For instance, Citizendium, a new electronic encyclopedia project, only accepts contributors who are registered with full curriculum vitae and proof of identity. The contributors must consent to obligatory disclosure of their Personal Identifiable Information (PII). However, users often prefer to be anonymous or pseudonymous when contributing contents or commenting on other contributions. In fact, it is crucial for protecting all our on-line privacy to be able to

* The extended version of this paper is available in the Cryptology ePrint Archive [6].
** Work done during internship at IBM Research - Zurich.

interact with such wide on-line communities in an anonymous or pseudonymous way. Moreover, pseudonymous interactions generally seem to guarantee higher quality of contributions.

An additional mechanism for quality assurance is *distributed moderation or rating* as, for instance, used by Slashdot.org or Apple's App store for the iPhone. Distributed moderation is typically done by rating, tagging, and reviewing of contributions or, in other words, by adding meta-data of the user community itself. It seems that such systems can quickly and consistently separate high and low quality comments in an online conversation [18], but also that the quantity and quality of meta-data may not be sufficient in practice unless users are given sufficient incentives. The latter was also observed in an experiment made on the IBM Intranet as part of the PrimeLife project [22]. Incentives could be in the form of monetary payments (e.g., micro-payments or points that can be redeemed later for a book or CD), valuations such as gaining reputation (cf. eBay), or in the form of side-effects (e.g., as games with a purpose [2]).

In conclusion, we need an on-line collaboration system that, on the one hand, protects the privacy of the users and, on the other hand, enhances the quality by giving incentives for reviews and moderation. For the latter, we need of course to ensure that the privacy offered cannot be abused. For instance, it must not be possible that one and the same person provides the original contribution and then also does all the moderation and reviews. In the paper, we first investigate the requirements and then provide a system that offers maximal privacy to the users, and allows for providing incentives and the establishment of reputations. Our system is based on unlinkable pseudonyms, anonymous credentials and e-cash [19, 4, 7, 8].

Contributions. We specify the first privacy-friendly incentive system with strong privacy protection and accountability. The system not only covers incentives and reputation, but also separation of duties, role-based and attribute-based entitlement policies. We provide a cryptographic realization based on abstract interfaces with zero-knowledge proofs of knowledge, anonymous credential systems and anonymous e-cash as primitives. Our system can be instantiated in the Strong RSA (SRSA) as well as in the Elliptic Curve Cryptography (ECC) setting.

We have implemented our incentive system for Wikipedia using the Identity Mixer cryptographic library [17] based on the SRSA. It can be used with any other on-line collaboration platform. We intend to make our source code publicly available at [22].

2 Privacy-friendly Incentives

Wikipedia provides documents to its users contributed by members of the community. This user-generated content varies in quality and can be significantly improved by (expert) reviews and comments. As most scientists know, good reviews are time-consuming, that is, come at a cost. Even though community service out of idealism is a common trait in the Wikipedia community, incentive systems can improve the situation for contributors as well as for the contributed content. They aim at reimbursing the review or revision cost by awards, and at invigorating the review process.

Privacy-friendly incentives complement this fundamental goal with anonymity and privacy protection for all users. Therefore, they enable a double-blind peer review process and nurture fairness, impartiality, and rigor. Authors as well as the reviewers of documents can remain anonymous during the entire review process. Such a review process is believed to be essential for academic quality, even though it sometimes lacks in reviewer accountability. Our goal is to establish a cryptographic system that reaches high quality standards, while fulfilling the diverse requirements of the involved parties.

We formalize the incentive system as a collaborative document editing system, in which all revisions, reviews and comments are linked to one initial document P_0. We consider a document version history $\mathbb{P} = \{P_0, \ldots P_n\}$ as ordered sequence of revisions, reviews and comments associated with the P_0, where P_n denotes the most recent revision or review.

Principals. There are multiple parties interacting with a document P. We have a clearing house that hosts all documents and organizes the incentive system, in our case the wiki W component. The wiki has a community of users and each user U may act in different and multiple roles:

Reader U**:** A reader consumes a document P. Any reader may offer incentives to other users to improve the quality of a document by a review or a revision.

Author V**:** An author contributes an initial version or a revision of a document P.

Reviewer R**:** A reviewer contributes reviews and comments for a document P in exchange for receiving an incentive.

Editor E**:** An editor is a privileged user, who may approve or decline document revisions or reviews by authors and reviewers.

We introduce a bank B to exchange electronic incentives for real-world goods and awards. Users of wiki W can withdraw fresh incentive e-coins and deposit spent ones as part of our virtual incentive economy. Even though we allow a system with full anonymity, we require each user to register with a trusted identity issuer I to infuse accountability in the entire review and incentive process. Each user U obtains an identity certificate σ_{U} on its identity sk_{U} from issuer I. Our system works with multiple banks as well as multiple identity issuers, we focus on the single-bank/single-issuer case for simplicity. The identity of an honest user is never revealed by the incentive system, whereas the certified identity enforces separation of duties between authors and reviewers, and prevents double-spending attacks as well as vandalism.

Concepts. In a privacy-friendly incentive system, many anonymous users interact with a single document P. Incentives may be given before or after a contribution (revision or review). *Pre-contribution* incentives are offered to users to provide a contribution at all and it is independent from the contribution quality. For instance, a reader U can offer incentive e-coins for any reviewer R who is willing to contribute a review. *Post-contribution* incentives are offered after the contribution is made and may be dependent on the quality of the contribution. For instance, users can rate the quality of reviewer's contribution and offer reputation e-coins for his work.

In our model, a reader U explicitly withdraws incentives from a bank B. The reader U offers these *pre-contribution* incentives on the wiki W for improvements on

a document P. The wiki W acts as a clearing house and it is responsible for ensuring unlinkability by exchanging the spent incentives of reader U with bank B for fresh incentives. Once a reviewer R decides to contribute a review P', he submits the review to the wiki W for inspection by an editor E. Once the editor E approves the review, the reviewer R can obtain the incentives from the wiki W. We leave community approval to the extensions in Sect. 5. As *post-contribution* incentives extension, the number of obtained incentives can be dependent on the review rating or the reviewer can obtain separate reputation e-coins to build a reputation credential.

Checks and Balances. The privacy-friendly incentive system provides anonymity to all users and balances this property with strong accountability safe-guards. In a fully anonymous system without such safe-guards, malicious users could attempt to manipulate reviews, sabotage other author's work or publish fabrications without accountability. Well known examples of checks and balances to counter those attacks are the separation of reviewer and author/editor, or the binding of reviews and documents to the contributor's true identity.

To achieve accountability as well as separation of duties between roles, we introduce a cryptographic domain pseudonym $N_{P,\mathsf{U}}$ for each user U that interacts with a document P. It is a function of the user's true identity sk_{U} and the document P while hiding sk_{U} computationally. Therefore, each entity interacting with document P has one unique pseudonym, which is independent from entity's role. Pseudonyms $N_{P,\mathsf{U}}$ and $N_{Q,\mathsf{U}}$ created for different documents P and Q are unlinkable.

3 Preliminaries

In this section we describe the abstract interfaces of the cryptographic primitives we employ, mostly following Bangerter et al. [4]. Our actual implementation uses and extends the Identity Mixer library [17] which offers the following primitives.

3.1 Commitment Schemes

A commitment scheme allows one to commit to a message m from some domain (typically $\mathbb{Z}_q$ for some prime q). The interface is as follows.

$C \leftarrow \mathsf{Commit}(m, r)$: Commit to message m via commitment C.
$\{0, 1\} \leftarrow \mathsf{VerifyCommit}(C, m, r)$: Verify commitment C belonging to message m.

The interface can be instantiated by the Pedersen commitment scheme [21] or the Integer commitment scheme by Damgård and Fujisaki [12]. For the Pedersen scheme, public parameters are a group G of prime order q, and generators $(g_0, \ldots, g_l)$. In order to commit to the values $(m_1, \ldots, m_l) \in \mathbb{Z}_q^l$, pick a random $r \in \mathbb{Z}_q$ and set

$$C \leftarrow \mathsf{Commit}((m_1, \ldots, m_l), r) = g_0^r \prod_{i=1}^{l} g_i^{m_i} \quad .$$

3.2 Zero-Knowledge Proofs and Σ-Protocols

When referring to the zero-knowledge proofs of knowledge of discrete logarithms and statements about them, we will follow the notation introduced by Camenisch and Stadler [11] and formally defined by Camenisch, Kiayias, and Yung [9].

For instance, $PK\{(a, b, c) : y = g^a h^b \wedge \tilde{y} = \tilde{g}^a \tilde{h}^c\}$ denotes a "*zero-knowledge Proof of Knowledge of integers* a*,* b *and* c *such that* $y = g^a h^b$ *and* $\tilde{y} = \tilde{g}^a \tilde{h}^c$ *holds,*" where $y, g, h, \tilde{y}, \tilde{g}$ and $\tilde{h}$ are elements of some groups $G = \langle g \rangle = \langle h \rangle$ and $\tilde{G} = \langle \tilde{g} \rangle = \langle \tilde{h} \rangle$. Following the approach of Bangerter et al., the PK notation accepts abstract predicates on input. For instance, $PK\{(m, r) : \mathsf{VerifyCommit}(C, m, r)\}$ denotes the proof of representation of a commitment. SPK denotes a signature proof of knowledge, that is a non-interactive transformation of a proof with the Fiat-Shamir Heuristic [14].

3.3 Signature Scheme for Anonymous Credentials

Bangerter et al. [4] formalize anonymous credential systems as an abstract signature interface. The signer is an issuer I with a key pair (sk_{I}, pk_{I}).

$(sk_{\mathsf{I}}, pk_{\mathsf{I}}) \leftarrow \mathsf{SetupSig}(\ell)$: Key generation for the issuer I.

$(\sigma)() \leftarrow \mathsf{HiddenSign}((C_1, \ldots, C_{l'}), (m_{l'+1}, \ldots m_l), r; pk_{\mathsf{I}})(sk_{\mathsf{I}})$: Issuer I signs hidden messages $(m_1, \ldots, m_{l'})$ in commitments $(C_1, \ldots, C_{l'})$ as well as known messages $(m_{l'+1}, \ldots, m_l)$. The user completes the signature σ with the commitment randomness r.

$\{0, 1\} \leftarrow \mathsf{VerifySig}(\sigma, (m_1, \ldots, m_l); pk_{\mathsf{I}})$: Predicate to verify a signature σ by issuer I on messages $(m_1, \ldots, m_l)$.

$\{0, 1\} \leftarrow \mathsf{VerifySigPred}(\sigma, (m_1, \ldots, m_l), \mathsf{AttrPredicate}; pk_{\mathsf{I}})$: Verifies additionally that the efficiently provable predicate $\mathsf{AttrPredicate}$ over the messages $(m_1, \ldots, m_l)$ is fulfilled.

This abstraction contains the following key points: First, it provides a $\mathsf{HiddenSign}()$ function that allows an issuer I to sign committed values $C_i = \mathsf{Commit}(m_i, r)$, $i \in \{1, \ldots, l'\}$ without knowledge of the hidden values m_i. Second, it provides a predicate $\mathsf{VerifySig}()$ that allows for a verification of signatures in zero-knowledge proofs of knowledge. Third, it offers an additional predicate $\mathsf{VerifySigPred}()$ to verify attribute statements over the attributes of signature σ in zero-knowledge proofs.

3.4 E-Coin Schemes

Our construction uses simple e-coins as a basic building block. We reference compact e-cash [8] for a formal set of definitions.

$(sk_{\mathsf{B}}, pk_{\mathsf{B}}) \leftarrow \mathsf{SetupBank}(\ell)$: Key generation for a bank B.

$(\sigma_\Psi, d_\Psi, s_\Psi)() \leftarrow \mathsf{Withdraw}(\sigma_U; sk_{\mathsf{U}}, pk_{\mathsf{B}})(sk_{\mathsf{B}}, pk_{\mathsf{I}})$: User U withdraws an unspent e-coin $(\sigma_\Psi, d_\Psi, s_\Psi)$ from a bank B, where σ_Ψ is the bank's signature on the e-coin, d_Ψ is a double-spending random element and s_Ψ is the e-coin serial number. The bank verifies U's identity sk_{U} certified by issuer I in signature σ_U.

$(T, R)(\Psi) \leftarrow \mathsf{Spend}(\sigma_U, (\sigma_\Psi, d_\Psi, s_\Psi); sk_\mathsf{U})(pk_\mathsf{B})$: User U spends an e-coin $(\sigma_\Psi, d_\Psi, s_\Psi)$ with a recipient while proving ownership of the e-coin with relation to its identity sk_U. The recipient outputs a spent e-coin $\Psi = (s_\Psi, R, T, \Phi)$, where R is a challenge from the recipient, T is a function of $(sk_\mathsf{U}, R, d_\Psi)$ such that $T \leftarrow g_B^{sk_\mathsf{U}\mathrm{R}} g_B^{d_\Psi}$ is computed by the user U and Φ is the proof transcript.

$()(\Psi) \leftarrow \mathsf{Deposit}(\Psi)()$: Sends a spent e-coin Ψ to the bank B.

$(pk_\mathsf{U}, \Pi) \leftarrow \mathsf{Identify}(\Psi_1, \Psi_2)$: Bank B runs Identify() on two spent e-coins Ψ_1 and Ψ_2 to identify the double-spending perpetrator U. It outputs U's public key pk_U and the double-spending proof Π.

$\{0, 1\} \leftarrow \mathsf{VerifyGuilt}(pk_\mathsf{U}, \Pi)$: The double-spending case (pk_U, Π) is publicly verifiable by VerifyGuilt().

Let us recall the core properties of an e-coin scheme:

Correctness. The Withdraw() and Spend() operations terminate successfully with honest participants. An honest recipient accepts an e-coin from a successful Spend().

Balance. No more e-coins can be spent than withdrawn.

Identification of Double-Spenders. Suppose bank B is honest. Let us consider U_1 and U_2 being honest users, each of them receiving the same e-coin during an execution of the Spend() protocol with an adversary, say $\Psi_1 = (s_\Psi, R_1, T_1, \Phi_1)$ and $\Psi_2 = (s_\Psi, R_2, T_2, \Phi_2)$. Then the adversary can be identified by the double-spending detection Identify() with overwhelming probability.

Public Key Recovery. The double-spending detection identifies a perpetrator U by outputting pk_U. We do not require full tracing as proposed in e-cash schemes.

Exculpability. Guilt in double-spending is publicly verifiable.

4 Core Incentive System

We define and analyze the core incentive system based on the preliminaries explained in Sect. 3. We start with definition of the service interface, continue with specification of the security requirements and realization of the system, and we conclude this section with security analysis.

4.1 Service Interface

$()(N_{\mathsf{B,U}}) \leftarrow \mathsf{Register}(\sigma_\mathsf{U}; sk_\mathsf{U}, pk_\mathsf{B})(pk_\mathsf{I})$: A user U registers at bank B anonymously while establishing a bank-specific domain pseudonym $N_{\mathsf{B,U}}$ for future transactions.

$(\sigma_\Psi, d_\Psi, s_\Psi)(N_{\mathsf{B,U}}) \leftarrow \mathsf{WithdrawIncentive}(\sigma_\mathsf{U}; sk_\mathsf{U}, pk_\mathsf{B})(sk_\mathsf{B}, pk_\mathsf{I})$: A reader U withdraws incentive e-coin from bank B. Reader U outputs a triple of e-coin signature σ_Ψ, double-spend random element d_Ψ and e-coin serial number s_Ψ. Bank B outputs the reader's domain pseudonym $N_{\mathsf{B,U}}$.

$()(\Psi, N_{P,\mathsf{U}}) \leftarrow \mathsf{SubmitOffer}(\sigma_\mathsf{U}, (\sigma_\Psi, d_\Psi, s_\Psi), P; sk_\mathsf{U})(pk_\mathsf{B}, pk_\mathsf{I})$: A reader U submits an incentive offer to wiki W, the clearing house. Wiki W outputs a spent e-coin Ψ and the reader's domain pseudonym $N_{P,\mathsf{U}}$ for document P. SubmitOffer guarantees reader's proof of possession of the e-coin.

$()(N_{P,\mathsf{R}}) \leftarrow \mathsf{ProposeReview}(\sigma_{\mathsf{R}}, P, P'; sk_{\mathsf{R}}, pk_{\mathsf{I}})(\mathsf{Review}_P; pk_{\mathsf{I}})$: A reviewer R proposes a review P' for document P anonymously at wiki W. The wiki W outputs the reviewer's domain pseudonym $N_{P,\mathsf{R}}$ for document P. It ensures that reviewer R fulfills the entitlement and qualification predicate Review_P and the separation of duties with the author.

$()(N_{P,\mathsf{E}}) \leftarrow \mathsf{EvaluateReview}(\sigma_{\mathsf{E}}, P, P', result; sk_{\mathsf{E}}, pk_{\mathsf{I}})(pk_{\mathsf{I}})$: An editor E rates the review P' for document P with rate $result$. The value $result$ determines approval or rejection. Wiki W enforces the separation of duties.

$(\Psi)(N_{P,\mathsf{R}}) \leftarrow \mathsf{SubmitReview}(\sigma_{\mathsf{R}}, P, P'; sk_{\mathsf{R}}, pk_{\mathsf{B}}, pk_{\mathsf{I}})(\sigma_{\mathsf{W}}, (\sigma_\Psi, d_\Psi, s_\Psi); sk_{\mathsf{W}}, pk_{\mathsf{I}})$: A reviewer R submits an approved review P' to wiki W and obtains the reward incentive e-coin Ψ in return. The domain pseudonym $N_{P,\mathsf{R}}$ links the transactions.

$()(\Psi) \leftarrow \mathsf{DepositIncentive}(\Psi)()$: A spent e-coin Ψ is sent to bank B.

$(\sigma'_\Psi, d'_\Psi, s'_\Psi)(\Psi) \leftarrow \mathsf{ExchangeIncentive}(\Psi, \sigma_{\mathsf{W}}; sk_{\mathsf{W}}, pk_{\mathsf{B}})(sk_{\mathsf{B}}, pk_{\mathsf{I}})$: Wiki W deposits a spent e-coin Ψ at bank B in exchange for a fresh e-coin $(\sigma'_\Psi, d'_\Psi, s'_\Psi)$. The bank B receives the deposited e-coin Ψ and may run $\mathsf{Identify}()$ to reveal double-spender.

4.2 Requirements

First, to ensure rigorous and impartial reviews, authors, reviewers and editors must benefit from a strong privacy protection. The parties need to be anonymous and their transactions unlinkable between multiple documents.

Second, we consider multiple access control properties. The system must support roles and attributes to qualify reviews. We also allow the certification of a reviewer profession and expertise, which increases trust in reviews and may entitle to claim a larger incentive for an editing task. In addition, the roles of different parties in a review process must be clearly separated. The most common example is that an author may not review and judge her own article.

Third, the system must hold users accountable for their actions to discourage vandalism and fraud. This involves a certification of users' identities, be it by the Wikipedia system itself or trusted third parties, such as government-supported electronic identification issuers. The system supports identity escrow by standard means, for instance, by verifiably encrypting user's true identity to a trusted anonymity revocation authority.

Incentive Security. *Correctness*: The operations terminate successfully with honest participants. *Balance*: No more incentive e-coins can be given than have been withdrawn. *Public Key Recovery*: An adversary can be identified by the double-spending detection with overwhelming probability, the perpetrator U is identified by outputting pk_{U}. *Exculpability*: Double-spending guilt is publicly verifiable.

Anonymity. The users of the Wikipedia system can be completely anonymous. Users shall be linked only to specific articles by domain pseudonyms.

Unlinkability. Different transactions within a review process, as well as transactions of the entire Wikipedia system are unlinkable. Unlinkability of the underlying technology is orthogonal to this claim (IP addresses, cookies, etc.)

Role- and Attribute-based Entitlement. The system allows role-based (RBAC) and attribute-based access control (ABAC) based on certified identities.

Separation of Duties. The system enforces a separation of conflicting duties (SoD). In particular, an author cannot review or rate her own article.

Accountability. The system holds users accountable for their actions by three means: (i) *Identity Certification*: The users' true identities and roles are certified in anonymous credentials by trusted issuers. (ii) *Master Key Consistency*: All credentials and transactions of a user U are bound to the same identity/master key sk_{U}. (iii) *Identity Escrow*: We allow a trusted third party to revoke the anonymity of users.

4.3 Realization

We realize the incentive service interface with the abstract primitives from Sect. 3. We focus on protocol diagrams for the complex interactions.

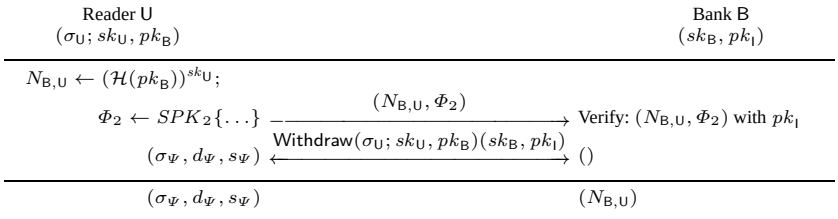
$\mathsf{Register}(\sigma_{\mathsf{U}}; sk_{\mathsf{U}}, pk_{\mathsf{B}})()$: We require each user U to register at bank B and to establish a bank-specific domain pseudonym $N_{\mathsf{B,U}}$ in the course of the action. User U proves knowledge of representation of the domain pseudonym $N_{\mathsf{B,U}}$ in SPK_1:

$$SPK_1\{(\sigma_{\mathsf{U}}, sk_{\mathsf{U}}, m_1, \ldots, m_l) : \\ \mathsf{VerifySig}(\sigma_{\mathsf{U}}, (m_1, \ldots, m_l); pk_{\mathsf{I}}) \wedge N_{\mathsf{B,U}} = (\mathcal{H}(pk_{\mathsf{B}}))^{sk_{\mathsf{U}}}\};$$

$\mathsf{WithdrawIncentive}(\sigma_{\mathsf{U}}; sk_{\mathsf{U}}, pk_{\mathsf{B}})(sk_{\mathsf{B}}, pk_{\mathsf{I}})$: We require a reader U withdrawing an incentive e-coin to prove her pseudonym $N_{\mathsf{B,U}}$ prior to the e-coin withdrawal with SPK_2:

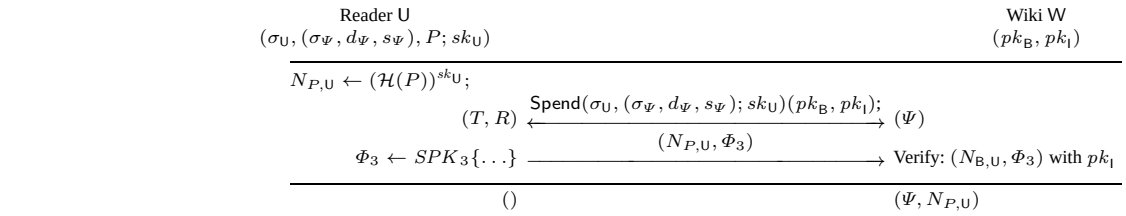
$$SPK_2\{(\sigma_{\mathsf{U}}, sk_{\mathsf{U}}, m_1, \ldots, m_l) : \\ \mathsf{VerifySig}(\sigma_{\mathsf{U}}, (m_1, \ldots, m_l); pk_{\mathsf{I}}) \wedge N_{\mathsf{B,U}} = (\mathcal{H}(pk_{\mathsf{B}}))^{sk_{\mathsf{U}}}\};$$

After the reader U successfully logged in as $N_{\mathsf{B,U}}$, it engages in a $\mathsf{Withdraw}()$ operation with the bank, to obtain the incentive e-coins.



$\mathsf{SubmitOffer}(\sigma_{\mathsf{U}}, (\sigma_\Psi, d_\Psi, s_\Psi), P; sk_{\mathsf{U}})(pk_{\mathsf{B}})$: To submit an offer, a reader U spends an incentive e-coin with the wiki W and proves knowledge of representation of his domain pseudonym $N_{P,\mathsf{U}}$ for a document P by SPK_3:

$$SPK_3\{(sk_{\mathsf{U}}, d_\Psi, R) : \qquad T = g_B^{sk_{\mathsf{U}}\mathrm{R}} g_B^{d_\Psi} \wedge N_{P,\mathsf{U}} = (\mathcal{H}(P))^{sk_{\mathsf{U}}}\};$$



$\mathsf{ProposeReview}(\sigma_{\mathsf{R}}, P, P'; sk_{\mathsf{R}}, pk_{\mathsf{I}})(\mathsf{Review}_P; pk_{\mathsf{I}})$: A reviewer R may propose a review P' for document P by proving knowledge of representation of domain pseudonym $N_{P,\mathsf{R}}$ to wiki W. SPK_4 proves in addition that the reviewer R certificate σ_{R} fulfills the predicate Review_P, e.g., that reviewer R is a doctor:

$$SPK_4\{(\sigma_{\mathsf{R}}, sk_{\mathsf{R}}, m_1, \ldots, m_l) : \mathsf{VerifySigPred}(\sigma_{\mathsf{R}}, (m_1, \ldots, m_l), \mathsf{Review}_P; pk_{\mathsf{I}}) \wedge N_{P,\mathsf{R}} = (\mathcal{H}(P))^{sk_{\mathsf{R}}}\};$$

The wiki W verifies that the reviewer R is different from the author of the document P by comparing their domain pseudonyms.

$\mathsf{EvaluateReview}(\sigma_{\mathsf{E}}, P, P', result; sk_{\mathsf{E}}, pk_{\mathsf{I}})(pk_{\mathsf{I}})$: An editor E can evaluate a review P' after having proven knowledge of representation of his domain pseudonym $N_{P,\mathsf{E}}$ in SPK_5:

$$SPK_5\{(\sigma_{\mathsf{E}}, sk_{\mathsf{E}}, m_1, \ldots, m_l) : \mathsf{VerifySig}(\sigma_{\mathsf{E}}, (m_1, \ldots, m_l); pk_{\mathsf{I}}) \wedge N_{P,\mathsf{E}} = (\mathcal{H}(P))^{sk_{\mathsf{E}}}\}(result);$$

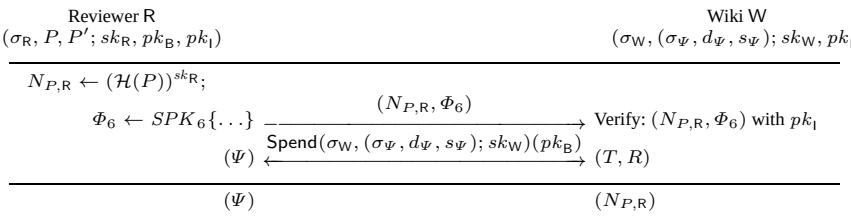
The wiki W verifies that editor E and reviewer R are different by comparing their domain pseudonyms.

$\mathsf{SubmitReview}(\sigma_{\mathsf{R}}, P, P'; sk_{\mathsf{R}}, pk_{\mathsf{B}}, pk_{\mathsf{I}})(\sigma_{\mathsf{W}}, (\sigma_\Psi, d_\Psi, s_\Psi); sk_{\mathsf{W}}, pk_{\mathsf{I}})$: When a reviewer R submits an approved review P', she needs to prove knowledge of representation of her domain pseudonym $N_{P,\mathsf{R}}$ first to link the transaction to the previous ones:

$$SPK_6\{(\sigma_{\mathsf{R}}, sk_{\mathsf{R}}, m_1, \ldots, m_l) : \mathsf{VerifySig}(\sigma_{\mathsf{R}}, (m_1, \ldots, m_l); pk_{\mathsf{I}}) \wedge N_{P,\mathsf{R}} = (\mathcal{H}(P))^{sk_{\mathsf{R}}}\};$$

The wiki W only engages in the $\mathsf{Spend}()$ protocol with reviewer R after successful proof. The reviewer R obtains an incentive e-coin Ψ and can subsequently deposit it at

the bank B.



4.4 Security Analysis

Incentive Security. The balance property of the e-cash system directly transfers to the incentive balance of our construction. The e-cash system's Identify() and VerifyGuilt() operations on user public keys enforce *Balance* and *Exculpability* properties.

Anonymity and Unlinkability. We based our construction on anonymous credentials as root identity. Throughout the system's transactions, users only prove knowledge of representation of their domain pseudonyms and their actual identities are kept confidential, except in the case of double-spending.

The cross-document unlinkability is maintained because the domain pseudonyms are uniformly distributed random group elements under the assumption of the random oracle model (ROM). The decision whether two keys, x and y, are equal given $(\mathcal{H}(P))^x, \mathcal{H}(Q))^y, P, Q)$ is hard under the Decision Diffie-Hellman (DDH) assumption in the ROM. We break the linking of non-transferable e-coins by an exchange between clearing house and bank via the ExchangeIncentive() operation.

Role- and Attribute-based Entitlement. We use the certified attributes in user's identity credential σ_{U} as flexible entitlement mechanism. Our system supports RBAC by certified role attributes, as well as ABAC by selective disclosure of further attributes. We employ this technique in the ProposeReview() operation: a reviewer R proves that her identity credential σ_{R} fulfills a review predicate Review_P. It applies to all of the proofs of representation of domain pseudonyms.

Separation of Duties. The separation of duties is enforced by proofs over domain pseudonyms. We realize *subject-based* separation of duties, the editor of the document is different from the author, by an inequality check of the domain pseudonyms. We realize *role-based* separation of duties, the document can only be confirmed under four-eyes principle, where one user has the role of a *Clerk* and another has the role of a *Manager*, by signature proofs of knowledge of roles and attributes associated with a domain pseudonym. The probability that two keys, x and y, for which $(\mathcal{H}(P))^x, \mathcal{H}(P))^y)$ collide is negligible as both are uniformly distributed random group elements given the ROM.

Accountability. We get *Identity Certification* and *Master Key Consistency* properties by design of the anonymous credential system. We achieve the *Identity Escrow* property by including a Verifiable Encryption (e.g., Camenisch and Shoup [10]) of a user's true identity sk_{U} towards a trusted third party. We consider this procedure a standard technique and do not elaborate on it.

5 Functional Extensions and Future Work

Although we implemented the core incentive system as elaborated in Sect. 6, we have not yet realized certain extension ideas. The following explains possible extensions for rating reviews on contributions and building reviewer's reputation.

Rating Reviews. In the presented system, the wiki W is responsible for checking the quality of the reviews and, if it finds the quality sufficient, for releasing spent incentive e-coins to the reviewer R. Alternatively, one could let the wiki community rate these reviews and have the wiki W only release the e-coins to the reviewer R, if the submitted review obtained a sufficiently high ranking. The mechanism would proceed as follows: Users (raters) sign their rating of the review with their domain pseudonym. The wiki W collects these ratings, checks that the domain pseudonyms of the raters and of the reviewer are different (separation of duties), as well as that each domain pseudonym of a rater only occurs once (one-time rating).

If there are several reviews, the offered e-coins can be distributed to different reviewers in proportion of the reviews ranking. This approach encourages reviewers to provide quality reviews in order to gain a high ranking and collect most of the e-coins. At the same time, it prevents reviewers from collecting all of the e-coins for poor quality reviews.

Reviewer Reputation. Rating of a review provides feedback on the quality of the review. This naturally lends itself to be used for an (anonymous) reputation system. Thus, the wiki W could issue reputation credentials (points) to reviewers and authors based on the quality of the reviews and articles. Articles could be ranked by users similarly to the reviews as described above.

More precisely, a reputation system can be implemented as follows. In addition to earned incentives, the wiki W could also issue an anonymous one-time credential to the user U (reviewer or author) according to the received average rating. This credential can be realized with the e-coin scheme, where the rating is encoded in the denomination. One either uses a different bank public key for each denomination or one extends the e-cash scheme to include denomination as an e-coin attribute. These *reputation e-coins* can then be gathered by the author or reviewers.

Let us assume some reputation authority that issues credentials which state user's reputation. Users can then exchange the reputation e-coins with the reputation authority against an updated reputation credential without this transaction being linkable to the corresponding article/review. The one-time spending property of the e-coin will ensure that each rating can be used only once. Depending on how the reputation is computed,

the rating e-coin and the old reputation credential cannot be exchanged directly for a new reputation credential, but the user U might need to have a pseudonymous account where he can deposit all the different ratings and then get an updated reputation credential issued once this computation is done. We leave a detailed discussion to the extended version of this paper.

6 Example Application

Wikipedia is a large-scale online encyclopedia project that at this time has grown to $\sim 3.35 \cdot 10^6$ articles in the English version and $\sim 9.25 \cdot 10^6$ articles in total [27], and that is beginning to rival more established compendiums of human knowledge [16]. Its software platform, MediaWiki, allows anybody with Internet access to read and edit shared articles. The most important criteria in Wikipedia's search for new quality assessment methods are the immediacy typical for social media as well as accuracy, which can be challenging [24].

The German Wikipedia chapter has deployed one such process in the form of the MediaWiki extension `FlaggedRevs` [13]. It allows eligible users, e.g., those who have earned editor role by being active community member for a certain duration, to review articles. In one configuration, review criteria include levels of accuracy, depth, and readability, and the review status is prominently displayed along with each article as important quality indication. According to the `FlaggedRevs` 2008 report [15], approximately 90.8% of the German Wikipedia articles have been reviewed at least once, even though, mostly by small pockets of active (expert) contributors.

Our example application functions as an extension to the `FlaggedRevs` extension and is registered as add-on PHP functionality in the MediaWiki. We register several incentive handling functions at its main code entry points. If both extensions, `FlaggedRevs` and `Mango` are installed, users can offer incentives when they want certain articles to be reviewed, and reviewers can earn these incentives by providing their expert insights through the revision process.

We report that we have implemented the presented incentive system architecture in the MediaWiki (see Fig. 2) as the `Mango` extension, as well as the cryptographic incentives system from Sect. 2 using the Identity Mixer library [17] based on the SRSA. The MediaWiki extension `Mango` contains the appropriate hooks to accommodate the cryptographic functions and serves as a glue between MediaWiki and the incentives system.

6.1 System Architecture

Static Design. Figure 1 shows the same entities as defined in Sect. 2 and concentrates on the software architecture of components that allow humans to participate in the scheme. High-level components have been realized as Java servlets.

Each user-facing component has an id, a password (corresponding to its MediaWiki account information) and a pair of cryptographic keys (for participating in the protocols). The bank B and the clearing W are special in that certain other components must

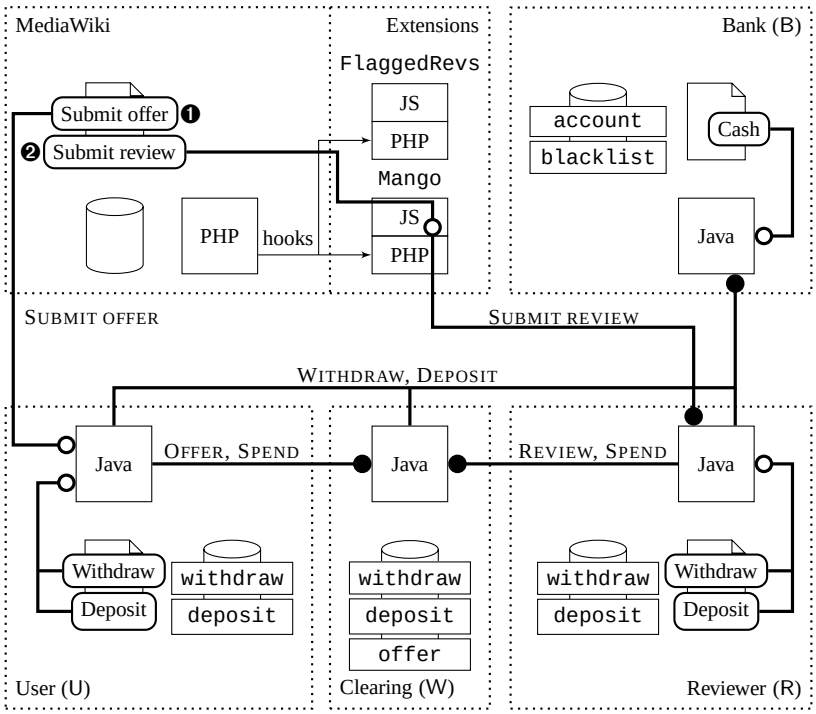


Fig. 1. System architecture.

have knowledge of these entities' public keys and network addresses in their configurations.

The components marked as *user* and *reviewer* correspond to U and R as defined in Sect. 2, and they receive communication at anonymous network addresses addr_{U} and addr_{R} respectively. The component marked as *clearing* corresponds to W, it receives communication at address addr_{W}. The clearing W functions as a front-end to MediaWiki and its extensions, thus linking its core logic (implemented in PHP and JavaScript by the MediaWiki conventions) to privacy-friendly incentives system (implemented in Java). The component marked *bank* corresponds to B and it receives communication at address addr_{B}.

All entity components maintain relational databases (cylinder shapes in Fig. 1) locally. The architecture does not assume that their private data is stored at any central location. For all indicated mappings, $a \mapsto b$ serves as a shorthand notation for the mapping from the set of all possible values for a to the set of all possible values for b.

Each bank B maintains two tables: The table `account` is a mapping $N_{\mathsf{B},\mathsf{U}} \mapsto (pk_{\mathsf{U}}, n)$, where $N_{\mathsf{B},\mathsf{U}}$ is a domain pseudonym computed by user U using address of the bank and n is the current balance of user U's account at the bank. The table `blacklist` is a mapping $N_{\mathsf{B},\mathsf{U}} \mapsto \{\, x_j \,\}_j$, where x_j are textual log entries pertaining to past double-spending behavior by user U, including the proof of double spending which can be verified by other parties.

Each user U maintains two tables: The table `withdraw` is a set of e-coins $\{ \Psi_i \}_i$ that were withdrawn from a bank B and have not yet been spent. The table `deposit` is a set of spent e-coins $\{ \Omega_i \}_i$ that have been received from another party, but have not yet been deposited at a bank B.

The clearing W is an extension of the user component and maintains one additional table: The table `offer` is a mapping $P \times N_{P,\mathsf{U}} \mapsto (n)$ where $N_{P,\mathsf{U}}$ is the reader U's domain pseudonym for article P and n is the number of e-coins offered for a review of the article P.

Dynamic Design. We will now explain the dynamic aspects of the system by following two representative use cases. The WITHDRAW, SPEND and DEPOSIT serve as high-level protocols calling the core incentive system's interface introduced in Sect. 4. In the first use case a user U offers privacy-friendly incentive points for an article review. The flow for this use case starts at the points marked ❶ in Fig. 1 (a user U presses the "Submit offer" button). Our walk-through assumes that an eligible user U has already logged into MediaWiki and that the system is in the right state. We present an elaborate version of this flow in the extended version of this paper.

Step 1.1 To offer e-coins for a non-stable (not reviewed) article, user U fills in an HTML form (see Fig. 2(a)) and presses the "Submit offer" button.

Step 1.2 The web browser submits user's request to the user U component, which runs locally on the user's machine as a Java servlet. The component U checks its `withdraw` table if sufficient number of unspent e-coins is available. In case of insufficient number of unspent e-coins, the component U will contact the bank B with user's consent, withdraw additional e-coins executing the WITHDRAW protocol and continue with submitting the offer.

Step 1.3 The user U and the clearing W components interact in the SUBMIT OFFER request to spend an incentive e-coin.

Step 1.4 The clearing W receives the e-coins, deposits them to the bank B, withdraws fresh unspent e-coins and stores them in its table `withdraw`. Identification of the article to which the offer pertains along with the number of e-coins offered are stored in the table `offer`.

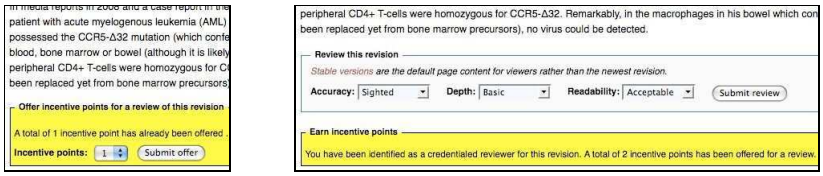
In the second use case a reviewer R receives privacy-friendly incentive points after conducting a review. The flow for this use case starts at the points marked ❷ in Fig. 1 (a reviewer R presses the button "Submit review"). Our walk-through assumes that an eligible reviewer R has already logged into MediaWiki and that the system is in the right state.

Step 2.1 A reviewer R chooses to submit a review. To do so, she fills in an HTML form (see Fig. 2(b)) and presses the "Submit review" button.

Step 2.2 This results in contacting the Java servlet of the reviewer R and an invocation of the clearing W involving both JavaScript and PHP.

Step 2.3 The clearing W looks up the entry in the table `offer`. On success, it spends the incentive e-coins with reviewer R using the SPEND protocol and it deletes the corresponding entries from `offer`.

Step 2.4 The reviewer R receives the spent e-coins and interacts with the bank B to exchange the e-coins for unspent ones (by executing DEPOSIT, then WITHDRAW) and stores the fresh e-coins in its table `withdraw`.



(a) Offering incentives. (b) Earning incentives.

Fig. 2. Our privacy-friendly incentives realization in use.

6.2 Anonymous and Pseudonymous Use

We note that MediaWiki already supports *pseudonymous use*. While this already affords relatively good privacy properties, our privacy-friendly incentive system from Sect. 2 can be extended to allow fully anonymous access to Wikipedia. In order to achieve this, the access control of MediaWiki needs to be adjusted to use domain pseudonym of a user and a proof that the user registered properly, the MediaWiki extension from [22] can be used to achieve this.We discuss the actual linkability in the MediaWiki system in the extended version of this paper.

7 Related work

Incentives are useful to create reputation systems. Steinbrecher studied privacy-protecting reputation systems [23] using pseudonyms. In such pseudonymous solutions, the transactions that are taken into account to build reputation can all be linked together. Therefore, many authors have claimed that achieving privacy in reputation systems is impossible [20]. In contrast, in our scheme one can build reputation from different transactions without these being linkable.

Adler and de Alfaro [1] proposed an orthogonal content-driven trust extension for MediaWiki, called WikiTrust. They focus on the analysis of a document's author, her reputation, origin, and trust, whereas our system considers the users' interactions in a double-blind review system. Lysyanskaya and co-authors [5] proposed and implemented an incentive system based on plain e-cash and fair exchange or file sharing applications. Their work focuses on the (fair) exchange of token and digital items. Androulaki and co-authors [3] proposed a reputation scheme for a pseudonymous peer-to-peer network. Their scheme uses e-cash to realize reputation points offered after a transaction is executed between network participants.

In contrast, we are interested in the (anonymous) relationships of the parties using e-cash to realize incentive points to enhance the quality of content. Furthermore, we propose possible extensions to our system to realize reputations of participating parties.

8 Conclusion

This paper has introduced novel concept of a privacy-friendly incentive system to rate user-generated content. We have proposed the first realization of such a system that draws on ideas from e-cash and anonymous credentials. The presented solution is privacy-friendly both from a theoretical and an applied perspective. In addition, we have contributed a practical architecture that integrates well with the open-source collaboration platform MediaWiki. To this end, we have extended MediaWiki for semi-anonymous use in a prototype environment, and designed the architecture such that it can support anonymous use by later adding anonymous credentials for authentication.

We report that we have implemented the cryptographic incentive system on top of the Identity Mixer library [7, 17]. We have realized a MediaWiki extension `Mango` with appropriate hooks for the cryptographic protocols and we integrated both results into MediaWiki, that is, hooking the implementation of the cryptographic protocols into the MediaWiki plug-ins.

We believe that providing such an incentive system nurtures high-quality content on electronic collaboration platforms by vigorous user interaction and rigorous double-blind reviews. We hope for a raise in quality and trustworthiness of user-generated content, in particular, if earned incentive points can be exchanged (at a suitable exchange rate) for real goods, such as CDs or vouchers.

Finally, note that our solution can be extended in multiple ways, most prominently through: (i) multifaceted incentives, (ii) transferable e-cash, (iii) identity escrow, and (iv) complex roles and policies.

Acknowledgment

This work has been funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 216483.

References

1. Adler, B.T., de Alfaro, L.: A content-driven reputation system for the Wikipedia. In: Proceedings of the 16th International Conference on World Wide Web (2007). pp. 261–270. ACM Press, New York, NY, USA (2007)
2. von Ahn, L.: Games with a Purpose. IEEE Computer Magazine (Jun 2006)
3. Androulaki, E., Choi, S.G., Bellovin, S.M., Malkin, T.: Reputation systems for anonymous networks. In: Privacy Enhancing Technologies. pp. 202–218 (2008)
4. Bangerter, E., Camenisch, J., Lysyanskaya, A.: A cryptographic framework for the controlled release of certified data. In: Twelfth International Workshop on Security Protocols. LNCS, Springer Verlag (2004)
5. Belenkiy, M., Chase, M., Erway, C.C., Jannotti, J., Küpçü, A., Lysyanskaya, A.: Incentivizing outsourced computation. In: NetEcon. pp. 85–90 (2008)
6. Camenisch, J., Groß, T., Hladky, P., Hoertnagl, C.: Privacy-friendly incentives and their application to Wikipedia (extended version). Cryptology ePrint Archive Report 2010/401, IACR (July 2010), `http://eprint.iacr.org/2010/401`

7. Camenisch, J., Herreweghen, E.V.: Design and implementation of the *idemix* anonymous credential system. In: Proc. 9th ACM Conference on Computer and Communications Security. acm press (2002)
8. Camenisch, J., Hohenberger, S., Lysyanskaya, A.: Compact E-cash. In: Cramer, R. (ed.) Advances in Cryptology — Eurocrypt 2005. LNCS, vol. 3494, pp. 302–321. Springer (2005)
9. Camenisch, J., Kiayias, A., Yung, M.: On the portability of generalized schnorr proofs. In: Joux, A. (ed.) Advances in Cryptology — EUROCRYPT 2009. LNCS, Springer Verlag (2009)
10. Camenisch, J., Shoup, V.: Practical verifiable encryption and decryption of discrete logarithms. In: Boneh, D. (ed.) Advances in Cryptology — CRYPTO 2003. LNCS, vol. 2729, pp. 126–144 (2003)
11. Camenisch, J., Stadler, M.: Efficient group signature schemes for large groups. In: Kaliski, B. (ed.) Advances in Cryptology — CRYPTO '97. LNCS, vol. 1296, pp. 410–424. Springer Verlag (1997)
12. Damgård, I., Fujisaki, E.: An integer commitment scheme based on groups with hidden order. In: Advances in Cryptology — ASIACRYPT 2002. LNCS, vol. 2501. Springer (2002)
13. Extension:FlaggedRevs, `http://www.mediawiki.org/wiki/Extension:FlaggedRevs`
14. Fiat, A., Shamir, A.: How to prove yourself: Practical solutions to identification and signature problems. In: Odlyzko, A.M. (ed.) Advances in Cryptology — CRYPTO '86. LNCS, vol. 263, pp. 186–194. Springer Verlag (1987)
15. Flaggedrevs report december 2008 (December 2008), `http://meta.wikimedia.org/wiki/FlaggedRevs_Report_December_2008`
16. Giles, J.: Internet encyclopaedias go head to head. Nature 438 (14 Dec 2005), `http://www.nature.com/nature/journal/v438/n7070/full/438900a.html`
17. IBM: Cryptographic protocols of the Identity Mixer library, v. 2.3. IBM Research Report RZ3730, IBM Research (2010), `http://domino.research.ibm.com/library/cyberdig.nsf/index.html`
18. Lampe, C., Resnick, P.: Slash(dot) and burn: Distributed moderation in a large online conversation space. In: ACM CHI 2004 Conference on Human Factors in Computing Systems (2004)
19. Lysyanskaya, A., Rivest, R., Sahai, A., Wolf, S.: Pseudonym systems. In: Heys, H., Adams, C. (eds.) Selected Areas in Cryptography. LNCS, vol. 1758. Springer Verlag (1999)
20. Pavlov, E., Rosenschein, J., Zvi: Supporting privacy in decentralized additive reputation systems. In: iTrust 2004 (2004)
21. Pedersen, T.P.: Non-interactive and information-theoretic secure verifiable secret sharing. In: Feigenbaum, J. (ed.) Advances in Cryptology – CRYPTO '91. LNCS, vol. 576, pp. 129–140. Springer Verlag (1992)
22. PrimeLife project, website. `www.primelife.eu`
23. Steinbrecher, S.: Design options for privacy-respecting reputation systems within centralised internet communities. In: SEC. pp. 123–134 (2006)
24. Viégas, F.B., Wattenberg, M., Dave, K.: Studying cooperation and conflict between authors with history flow visualizations. In: Proceedings of the 2004 Conference on Human Factors in Computing Systems, CHI 2004 (2004)
25. A little sleuthing unmasks writer of Wikipedia prank (December 2005), `http://www.nytimes.com/2005/12/11/business/media/11web.html`
26. Wikipedia to limit changes to articles on people (August 2009), `http://www.nytimes.com/2009/08/25/technology/internet/25wikipedia.html`
27. Wikipedia:Size comparisons (August 2010), `http://en.wikipedia.org/wiki/Wikipedia:Size_comparisons`