Alexandria

Matthew Fulton, Hardik Vala, Cameron Rout, Fernando Caudillo

Abstract

Alexandria is a decentralized protocol for facilitating blind academic peer review without a human intermediary. Authors request review of their manuscripts from sets of verified, credentialed reviewers, using a single-use, anonymous cryptographic identity. Any reviewer in the credentialed set may accept and perform the review by creating a single-use cryptographic identity and providing a ring signature, proving membership in the credentialed set without revealing individual identity. This author-reviewer matching, as well as all review communication is published in encrypted form to a public blockchain. Either party may reveal individual messages by publishing the message-specific encryption key. Selective message revelation combined with the use of ring signatures corresponding to credentialed rings allows any third-party to verify that a given manuscript was reviewed by an appropriately credentialed peer, without revealing the specific identity of that peer. At the conclusion of the review, the author may claim the review by publishing a signature from their known, public identity, proving ownership of the single-use identity created for the review.

Motivation

Peer review is vital to quality control in all scientific disciplines. All academic publications are subject to reviews by experts before being disseminated to the broader research community. Today, editors at scholarly journals assume the responsibility for organizing and orchestrating the peer review. This human intermediary is necessary in order to facilitate anonymous communication between author and reviewers, as well as provide future readers trusted assurance that published papers were expertly reviewed, without necessarily revealing reviewer identity.

However, the advent of decentralized distributed ledger technology, “blockchain”, combined with ring cryptography, offer an alternative. The Alexandria protocol leverages the blockchain to support academic peer review that doesn’t require any journal-human intermediary. Authors and reviewers can communicate over Alexandria while keeping their identities anonymous. This enables smaller/newer journals to prove credibility of their peer review process, and opens new pathways to publishing peer reviewed research.

The Alexandria Protocol

Review lifecycle

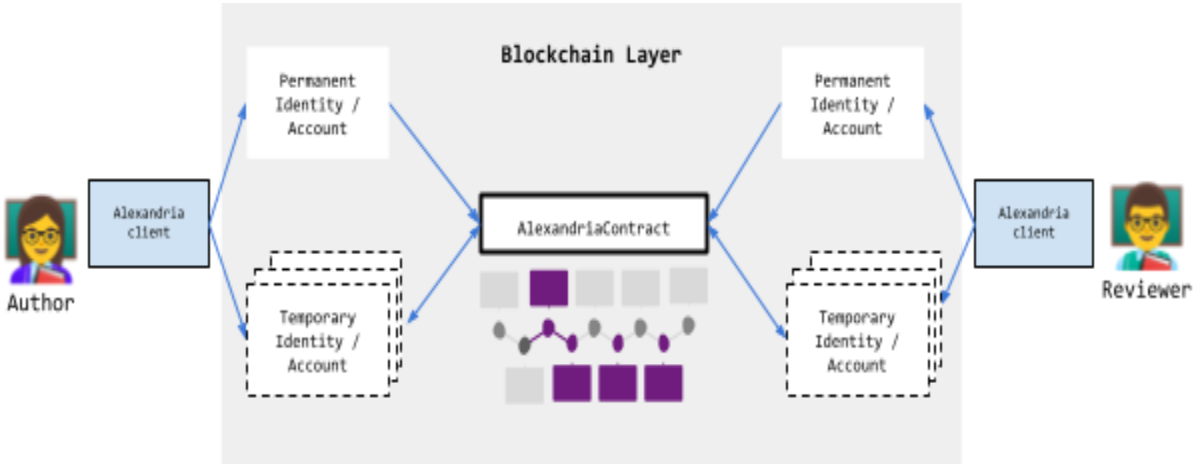
Alexandria models itself after the procedure by which double-blinded peer reviews are accomplished today, but with a cryptographic, rather than journal-human, intermediary. A review on Alexandria occurs as follows:

1) **Author submits application**: An author launches the review process by creating a new identity (account on chain) from which they submit a review request to the desired reviewer ring, as well as a link to the abstract of the manuscript to be reviewed. This request may be encrypted using the reviewer ring's public key, so only ring members can view the manuscript or abstract contents.
2) **Reviewer accepts a review**: Any member of the reviewer ring may accept the application by initiating the review. The reviewer first creates a new, anonymous identity from which they will conduct the review. They then accept the request by submitting a message consisting of a ring signature, proving their membership in the ring without revealing specific identity and a new public key, henceforth referred to as a *message encryption public key* (discussed in-depth below), distinct from the public key associated with the new identity, that the author should use to encrypt the next message.
3) **Author submits the manuscript:** The author uploads the manuscript to a content storage system fitting the parameters described below. The author then sends a link to the manuscript, encrypted with the reviewer-provided message encryption public key, along with a new message encryption public key for use in the reviewer's response.
4) **Reviewer sends review**: The reviewer evaluates the manuscript and uploads their feedback to the content storage system. The reviewer then sends a link to this review, their verdict, and a new message encryption key, encrypted with the most recent message encryption public key provided by the author. There are three possible verdicts:
    - APPROVE — Work should be published, optionally conditional on suggested revisions.
    - REJECT — Work should not be published, along with comments explaining why.
    - REVISE — Decision is pending answers to reviewer questions and suggested improvements, along with any other relevant comments.
5) **Author sends revision**: If the reviewer requests a revision (REVISE) before making their final decision, the author submits a new draft based on reviewer feedback, any responses to comments, and a new message encryption public key, encrypted with the latest, reviewer-provided message encryption public key.

6) **Author-Reviewer iteration**: Repeat steps 4-5 until the reviewer approves or rejects the manuscript.

Architecture

While not dogmatic in the specific choices, the Alexandria protocol does require components (eg, blockchain) with certain traits (eg, account-based vs UTXO).



Blockchain

Alexandria is designed to run on a blockchain with the following features:

- Supports Turing-complete code execution (smart contracts)
- Historical transactions are publicly verifiable
- Users have cryptographic identities
- Users may manage multiple unconnected identities

Suitable blockchains include Ethereum[1], Solana[2], layer 2 chains such as Polygon[3] or Stacks[4], etc. Unsuitable chains include Bitcoin[5] (does not provide Turing-complete code execution; layer 2s on top of Bitcoin would, however, be suitable).

Smart Contract

Alexandria is designed to be implemented as a smart contract on top of a blockchain with the properties described above.

---

[1] https://ethereum.org/en/whitepaper/
[2] https://solana.com/solana-whitepaper.pdf
[3] https://whitepaper.io/document/646/polygon-whitepaper
[4] https://stx.is/nakamoto
[5] Satoshi Nakamoto. Bitcoin: A Peer-to-Peer Electronic Cash System. Whitepaper, 2009.

Content Storage

Alexandria requires storage of review communication and content, such as manuscripts. This storage must have the following traits:

- Immutable — both content and address must not be changeable once set
- Ordered — which message came first?
- Sufficient space for communication (the memo field on Ironfish[6] would not, for example, be of sufficient size)

It is possible to implement Alexandria such that all communication (reviews, manuscripts, etc) is stored on the blockchain hosting the smart contract. This is, however, prohibitively expensive. This paper instead presumes the use of a system like IPFS[7]. Content on IPFS is immutable with the address being a hash of that content, allowing third-party verifiers to determine what was actually reviewed. All content is shared this way, through links embedded in transactions.

Alexandria Client

All participants use an Alexandria Client application, that runs locally and outside the blockchain, to interact with the Alexandria Protocol. Its purpose is to handle identity management and enhance user experience.

This client or clients could take the form of a website, mobile application, command line tool, or any other software which runs on a computer that may interface with the blockchain hosting the Alexandria smart contract.

Identities

There are two forms of cryptographic identities in Alexandria. Both forms correspond to identities on the hosting blockchain (often referred to as a “wallet”).

Public identities

Public keys associated with a real-world identity. These could be linked to ORCHID iDs[8], keybase.io keys, or simply a PK published on a researcher’s website. These are the identities registered with reviewer rings.

---

[6] https://ironfish.network/docs/whitepaper/
[7] https://ipfs.tech
[8] https://orcid.org/

Single-use identities

In order to preserve anonymity during the course of a review, authors and reviewers create new identities for each review. These are the identities from which all review communication occurs.

Authors and reviewers create these single-use identities on the hosting blockchain. They fund them via a crypto exchange (eg, Coinbase) or other fiat onramp (eg, MetaMask), which can be simplified using the Alexandria client. Note that exchanges often mandate users verify their individual identity to comply with local regulation. Users must entrust the exchanges to keep this sensitive information private.

Linking public and single-use identities

In order to claim a review, an author publishes a message signed with the private key of the single-use identity used in a review, from their known, public identity. See *Author Attribution*, below.

Reviewer rings

A reviewer ring is a set of public keys tied to known identities that all claim a credential relevant to peer review. An example reviewer ring could be “published physical chemistry researchers at accredited universities”.

Signature scheme

When a reviewer accepts a review request from a single-use identity, they provide a ring signature proving membership in the ring which the author requested the review from.

There are multiple viable signature schemes, which implementers may choose from. In fact, implementations may allow any of a whitelisted set of schemes, in order to allow migration to a new scheme in the face of deprecation of weakened cryptologic functions. Timestamping deprecations, while also including the scheme identifier in the original, timestamped signature message, would allow verifications of historical signatures using deprecated schemes, while providing a vehicle for scheme upgrades.

Any chosen scheme **must** be linkable (allow provability that two signatures came from the same group member) in order to combat certain attacks (see *Attacks*, below). All group signatures have this property inherently (the group administrator may reveal identity of a signer) as do some ring schemes.

Given the scarcity and associated cost of blockspace, the chosen scheme **should** produce short signatures that do not scale with ring size, as many ring signatures do.

The Alexandria protocol is designed with the assumption that, at the protocol level, all ring members have the same privileges. This rules out group signatures entirely.

This paper recommends the short, linkable ring signature scheme introduced by Tsang and Wei in 2005.

Ring versioning

Verifying a ring signature requires knowledge of the full set of public keys used to generate that signature. Given that membership can change over time, verifying a past signature (to verify the validity of a review) requires historical membership knowledge. Due to hosting ring membership on a public blockchain, full history is available, and any verifier can compute historical membership.

A race condition may arise such that a reviewer submits a ring signature while a membership change is in-process. In this case, the ring signature should be rejected, and the reviewer will be expected to resubmit, with an updated signature.

The meta ring

There is one special-purpose ring, henceforth referred to as *the meta ring*, whose sole purpose is to review the creation and modification of rings and their memberships. The criteria that meta ring reviewers use to evaluate these requests are clearly and strictly defined. The role of meta ring reviewers is primarily to verify confirmable facts, rather than make normative judgements.

Membership changes to the meta ring are conducted by two meta ring reviewers. If those reviewers disagree, a third reviewer makes the final decision.

Registration

New rings are created via the following following process:

1. A ring creator creates a proposal for a new ring, and requests review from the meta ring. This proposal will include the relevant research area, any additional criteria, initial member public keys, contact information, and proofs of credentials.
2. Like any other Alexandria review, a member of the meta ring performs the review, as described in the *Blind Reviews* section, below. In this case, the reviewer will be

responsible for reviewing all materials, confirming member identity, and confirming member interest in participating in the review ring.

3. Once approved, a ring is created on Alexandria

Ring management

Adding a new member

Adding a new member to a ring requires that new member to submit a proposal to join a ring, with sufficient proof of credentialing, to the meta ring. Meta ring reviewers then confirm claimed public identity, desire to participate in the ring, and proof of credentialing.

Removing a member

Any member may remove themselves from a ring at any time. No review is required.

There are several non-exclusive options for removing members other than yourself:

Option A: Anyone may submit a member removal request to the meta ring. Meta ring reviewers should only remove ring members who have lost relevant credentials or have out-of-date verifications of their timebound credentials (discussed below).

Option B: Implementations may choose to only remove ring members due to failures to renew before membership expires. In this case, it would not be necessary for the meta ring to review removal requests. These implementations should hook ring membership cleanups into other actions, such as cleaning up ring memberships whenever an author requests a review from a ring.

Option C: Implementations may provide a global cleanup() function, which anyone could call. Depending on the implementation environment, it is likely that calling this global function would require payment of blockchain fees, and thus require a reward or external incentive to do so.

Updating member verification

Some credentials may be timebound. A ring of “active” researchers may require publication in the field, in the preceding 12 months. Each ring member should provide updated proof for each timebound credential, within the period, by submitting a credential verification request to the meta ring.

Alexandria implementations should “deactivate” members with expired credentials, such that their public keys are not used to generate ring signatures, (thus preventing those individuals from performing reviews). This deactivation may take the form of an eviction of

out-of-compliance members, processed with each new review request or acceptance (when ring signatures must be verified).

Updating ring metadata

A change in credential requirements could result in an eviction in some number of members, and thus is not permitted. Instead, a ring wishing to change credential requirements should submit a proposal for a new ring.

Updates to non-credential metadata require a proposal, approved by the majority of ring members.

Deprecating a ring

Alexandria has no direct role in deprecating a ring. Alexandria relies on (a) all members taking the responsibility of removing themselves from rings they wish to no longer be a part of and (b) a client to have the ultimate responsibility of recommending only active rings by looking at review history on chain, looking at recent membership updates, or other activity. Note also that rings with timebound credentials will naturally self-deprecate when out-of-use.

Ring governance

Alexandria is designed such that all ring members have the exact same set of privileges. More complex ring governance is possible, but should happen outside of the Alexandria protocol.

Blind Peer Review

Alexandria mediates peer review as a trustworthy broker by inheriting the security properties of the underlying blockchain, and by relying on methodical bookkeeping with careful orchestration of interactions between multiple parties. The `AlexandriaContract` facilitates all communication, on an open blockchain, and depends on strong encryption to protect identities, and therefore unintended bias.

Single-use identities

As described above, reviewers and authors communicate from identities created specifically for the review. Reviewers prove membership in the given reviewer ring by including a ring signature with their initial review acceptance message. At the conclusion of a review, authors may “claim” a review by publishing a signature from their known, public identity, with the private key of the single-use identity used for this review.

Obscuring Content

Abstract and initial review request

Authors provide an abstract when requesting review from a ring. To support cases where the abstract should not be publicly available until after review and subsequent publication, the abstract should be encrypted using a key generated by the author. The linked document containing the encrypted abstract should also a list of pairs of the public keys of reviewers in the ring, and the public-key-encrypted abstract decryption key.

Alternatively, an implementation could simply encrypt the abstract once for each public key in the ring, and have authors bear the corresponding additional storage costs.

Review communication

Each message between an author and reviewer is public by the nature of public blockchains. Reviewers and authors use the following protocol to shield these messages from public view: each message, starting with the message where a reviewer declares their intent to review a submission, includes a single-use public key — a message encryption public key. When responding to a given message, the recipient uses the provided message encryption public key to encrypt the contents of their response, which includes its own message encryption public key from a freshly-generated key-pair. This protects the integrity of the review process while it's active. This allows selective publication of any review message, by either party. Critically, this enables authors to reveal an "approval" message from a reviewer, without revealing all communication which occurred during the review. This also allows either party to reveal the contents of any single message without revealing all messages exchanged during the review.

Triple-Blindness with Parallel Reviews

Authors can submit multiple review applications for the same article to a reviewer ring, to acquire multiple independent reviews. This is consistent with standard practice, where journals will have a group of reviewers assess the same submission, independently. Today, journal editors isolate reviewers from each other to avoid corrupting feedback - a third dimension of blinded-ness. In Alexandria, every review interaction, including the reviewer-submitted approval, is not public by default. This allows for *triple-blindness* across parallel reviews, by keeping review verdicts secret until the author chooses to reveal it, thus allowing for all parallel reviews to complete without knowledge of the verdicts of other reviewers.

The author or reviewer can choose to expose the contents of these interactions after the review is completed (see *Review Verification*, below), meaning it is possible for a reviewer or author to

break this third level of blindness. But clients of Alexandria will be able to determine when a review verdict (and thus final content) has been completed, and thus determine whether there were any subsequent verdicts that may have been biased. There is thus social pressure (and likely formalized rules in some publications) for authors to keep review results confidential until acquiring approvals from a sufficient number of reviewers.

Author Attribution

An author can claim a review by publishing a signature proving ownership of the private key associated with the single-use identity used for the review, from their public, known identity. (This is only possible after a review has concluded with an APPROVE / REJECT decision.)

Review Verification

To add transparency into a review, Alexandria allows the public release of the contents of any message by either (a) publishing the private key associated with the public key used to encrypt it, or (b) publishing the content of the message, unencrypted, such that any third party could use the associated public key to verify that the published content is in fact what was encrypted. Critically, authors use this mechanism to reveal any approval messages.

Blockchain Fee Payments

Many Blockchains, including Ethereum, impose fees for transacting on the network. Single-use accounts are funded as described above. Further, authors are required to fund anticipated reviewer transaction fees. Ideally, this would be done on a message-by-message basis, allowing for a pay-as-you-go system. Implementations may hold these fees in escrow, releasing them upon reviewer response, or may directly funnel fees from authors to reviewers. In the latter case, there should be controls preventing material overpayment of fees (bribes, effectively).

Timeouts

Either author or reviewer may stop responding to the review at any point. This leaves the other side hanging, potentially wasting the time invested. As such, Alexandria implementations should include a response deadline. Once that deadline passes, the responsive party should have the option to cancel the review, and receive any relevant fees and/or staked funds (see *Incentives*, below).

Incentives

Alexandria neither proposes any additional incentives beyond what already exist in academia, nor forbids them — an implementation may introduce a financial incentive, in the form of allowing reviewer rings to set a fee required for review, for example.

In the current system, reviewers are typically also authors in their own right. They conduct reviews both to contribute to their field as well as shape that field by evaluating manuscripts. Alexandria doesn't alter these incentives.

Implementations may consider using financial incentives to replace social incentives that exist with human-journal intermediaries. For example, an implementation may require reviewers to post funds when accepting review, which would be distributed to the author in the case where a reviewer fails to provide a review within deadline. Similarly, implementations may choose to distribute review fees to the reviewer if an author fails to respond within deadline.

Attacks

Denial of service attacks

A malicious blockchain user may spam a reviewer ring with a large number of review requests, from different anonymous identities. In order to prevent this, review requests require a payment large enough to prevent this, without being burdensome to potential authors. A fee of $O(10^1)$ dollars should be sufficient.

Hidden rejections: authors claiming only reviews that result in approvals

An author may request reviews repeatedly until receiving a sufficient number of approvals for publication, never revealing the rejections. While this is possible in theory, it is unlikely to occur in practice for social reasons. All members of a ring may view review requests (and associated abstracts or manuscripts), allowing for identification of papers with a large number of review requests. With wide knowledge of this form of attempted author abuse, reviewers who reject the paper have an incentive to reveal their rejections on the public chain, thus revealing a more complete picture of the peer evaluation of the paper.

Single reviewer, multiple reviews

The linkability of signatures in the chosen ring signature scheme prevents this.

Author <> Review coordination

Authors and reviewers may conspire to approve each others' papers without "real" review, by coordinating on when the author will submit the review request, such that the colluding reviewer can be ready to immediately accept that request. Collusion between researchers already occurs in non-blockchain contexts, and Alexandria does not propose additional barriers beyond what already exist. Primarily, for a paper to be considered credible, it needs reviews from multiple reviewers. Norms will develop over time as to the quantity of reviews required for a paper to be considered credible.

Leaks

Content leaking

Reviewers may leak a manuscript during a review, before the author is ready to publish. Due to the linkability of ring signatures, however, it is possible to compute the probability that a reviewer who accepts a review request has leaked in the past. By way of example: say there were 3 concurrent reviews during a leak. If one of those ring signatures is linkable to a future ring signature used to accept a future review, there is a ⅓ chance that person was the linker. The author could choose to stop the review, with only the abstract submitted with the review request, thus redirecting traffic away from malicious reviewers. Fees should be returned to authors in this scenario.

Single-use identity doxxing

Authors and reviewers use fiat onramps (Coinbase, MetaMask, etc) to fund single use identities used in review. The companies backing these onramps typically comply with regulation requiring knowledge of identity, and thus could reveal the public identities of single-use, anonymous identities used to conduct a review.

Those entities are unlikely to doxx authors/reviewers in this way, as authors/reviewers would simply move their business elsewhere.

Review fee farming

Review fees — whether nominal in order to prevent DoS attacks or material due to an implementation allowing for rings to set fees — provide an incentive for reviewers to accept and quickly approve any incoming reviews, in order to earn those fees for minimal work.

Alexandria mitigates this through the use of linkable ring signatures. Publications or any third party (including anyone reading the chain using any client) may compare the ring signature of reviewers of submitted papers to those with activity that appears to be fee-farming. Publications and readers may then discount reviews from signatures associated with fee-farming. Authors would then be incentivized to withhold manuscripts from reviewers who accept reviews with signatures linked to known farmers. Fees should be returned to authors in this scenario.

Discontinuing reviews from high rejection-rate reviewers

To avoid review fee farming by reviewers, authors may choose to not start a review after its acceptance by a reviewer, and receive the paid fee back. This opens the door to authors looking at the accept/reject rates of reviews performed under the authority of ring signatures linkable to the signature provided in accepting their current review request, and decide to only proceed with reviewers with high approval rates. Alexandria provides two mitigations:

1. Delay in fee rebates. A small fee should be held for a long time (or not returned at all) and a large fee should be held for enough time to discourage gaming the review system in the above-described way (months, maybe).
2. Social defenses. All ring members can see incoming review requests, and will thus be able to identify abstracts that are submitted an unusually high number of times, indicating potential author gaming. Ring members could then choose to publish this information, hurting author and paper credibility.

Exaggerated expertise: leveraging the credibility of inactive ring members

Ring signatures may be constructed from any list of public keys. Malicious Bob could create a ring signature, for example, proving a message was produced by one of Malicious Bob, Innocent Alice, or Vitalik Buterin. In this example, a recipient of the signature may infer that Malicious Bob is on even footing with Innocent Alice or Vitalik Buterin, or that one of them were the actual signers, even if neither had any knowledge of Malicious Bob.

In a review context, this allows weakly credentialed groups to include keys of those with stronger credentials, increasing the weight of reviews from that ring.

Alexandria mitigates this by requiring all members in the ring to verify association desire as part of the ring registration process. Rings with time-bound credentials would also implicitly require regular affirmations of association desired by all parties. Thus, if Innocent Alice and Vitalik Buterin affirm their desire to be part of Malicious Bob's ring, authors can assume that, at the very least, Vitalik and Alice support the ring's claimed credentials as a whole, and may also surmise that they provide some portion of reviews. Note that Alexandria does not provide a vehicle for requiring/enforcing individual ring member review participation. It is not necessary for the protocol to be effective, and therefore this and any other additional ring governance would have to occur outside of the protocol itself.