Callisto: A Cryptographic Approach To Detect Serial Predators Of Sexual Misconduct

Anjana Rajan Lucy Qin David Archer Dan Boneh Tancrède Lepoint Mayank Varia March 29, 2018

Abstract

Callisto, a non-profit that has created an online sexual assault reporting platform for college campuses, is expanding its work to combat sexual assault and harassment in other industries. In this new product, users will be invited to an online "matching escrow" that will detect repeat perpetrators and create pathways to support for victims. Users of this product will be able to enter the identity of their perpetrator into the escrow. This data can only be decrypted by the Callisto Options Counselor (a lawyer), when another user enters the identity of the same perpetrator. If the perpetrator identities match, both users will be put in touch independently with the Options Counselor, who will connect them to each other (if appropriate) and help them determine their best path towards justice. The client relationships with the Options Counselors are structured so that any client-counselor communications would be privileged. A combination of client-side encryption, encrypted communication channels, oblivious pseudo-random functions, key federation, and Shamir Secret Sharing keep data encrypted so that only the Callisto Options Counselor has access to user submitted data when a match is identified. In this paper, we present an informal risk management assessment, threat model, and cryptographic solution overview for our new product. A later paper will provide a formal security analysis and mathematical proofs of our cryptographic scheme.

1 The Problem of Sexual Assault and Harassment

An estimated 20% of women, 7% of men, and 24% of transgender and gender non-conforming students are sexually assaulted during their college career. Less than 10% of college assault survivors report to administrators, local police, campus security, or other authorities. Those who choose to report do so an average of 11 months after their assault, making it hard to conduct an investigation. These investigations are not challenging because perpetrators are unknown - in fact, 85% of college survivors know their assailant -but rather because investigators are not sure whether to believe that an assault actually took place. Only 6% of assaults reported to the police end with the assailant spending a single day in prison, meaning that over 99% of those perpetrators will not face serious consequences for their actions. Thus there is no effective deterrent to sexual assault in the United States. [4][3]

Meanwhile, an estimated 90% of college sexual assaults are committed by repeat perpetrators. These serial perpetrators assault an average of 6 times before they graduate from college. But with such a low reporting rate, it is fairly unlikely that even serial perpetrators will be reported, much less reported more than once. Therefore, investigators often have no knowledge of a pattern of behavior of the accused when trying to make a fair judgment on a case. Without clear evidence (which is hard to gather if a report is made months after an assault) or a pattern of behavior, authorities are often hesitant to assume liability

for taking action against an accused perpetrator. It is far more common for colleges to be sued for expelling an accused sexual assailant than for neglecting the safety and privacy rights of victims.

We might create a sea change if victims could learn that they are "not the only one". In a survey conducted by Callisto of over 200 college sexual assault survivors, a clear pattern emerged - for most victims, reporting an incident of sexual assault was not worth it, unless they knew they were not the only victim of that assailant. Learning of another victim of the same assailant dramatically increased victims' willingness to report, as well as their perceived likelihood of being believed if they reported. However, in our current environment, the only way they can learn of other victims is through the "whisper network", or if their perpetrator is identified in the press or on social media.

As the #MeToo movement has manifested, it has become clear that sexual assault and harassment, especially in the workplace, is prevalent across many industries. Incentives have shifted so that employers and professional sectors are now facing the same kind of pressure that colleges faced 4 years ago: pressure to take tangible action to address sexual misconduct. However, employers face many of the same issues as college investigators: delayed reports with little evidence other than the testimony of the victim, and with no good way to learn whether the accused exhibited a pattern of behavior. Victims in the workplace face the same stark reality as college survivors - that it is often not worth it to report unless you know that you are not the only one. The Equal Employment Opportunity Commission's 2016 report on workplace harassment found that almost one third of the approximately 90,000 complaints received by the EEOC in fiscal year 2015 included an allegation of workplace harassment [1]. Roughly three out of four employees who experienced harassment never talked about it with a supervisor, manager or union representative. Employees choose not to report or file a complaint because they fear disbelief, inaction by management, blame, or social or professional retaliation. While everyone can agree that the current equilibrium does not work, it remains in the best interest of both victims and authorities to continue the culture of silence. If nothing changes, once the #MeToo movement fades, victims will continue to not report, authorities will resume not taking action, and serial perpetrators will continue to assault and harass more victims.

Unfortunately, under-reporting by victims and associated non-accountability of perpetrators is not the only problem to be solved. As much as we believe that efforts to stop sexual assault and harassment would benefit from full disclosure and transparency, that same ideal works against victims. A victim's identity, details of incidents, and identities of perpetrators are all highly sensitive information. In the wrong hands, that information can be used to cause serious harm. Worse, it can be used to inhibit the reporting and follow-up so necessary to helping victims find justice. Perpetrators would certainly use the knowledge that they might be reported to intimidate, threaten, or take legal action against victims. More importantly, society often uses such information to damage victim or perpetrator reputations or wellbeing. Finally, and most importantly, each victim's right to choose their path to justice is paramount. Victims need a way to discover the paths open to them and find support on those paths while retaining their privacy and personal security.

2 Solution Overview

Callisto approaches the problem of hesitance in reporting assault by using the mathematical tool of *game theory*: a way of modeling situations of conflict among parties [2]. In game theory terms, there is a first-mover disadvantage with high consequences for the victim when accusing a perpetrator. That disadvantage comes from the disclosure and resulting exposure of the victim, opening the victim up to consequences (countermoves

in game theory) of retaliation, disbelief by authorities, reputation damage, and stigma. Callisto's solution aims to eliminate first-mover disadvantage by allowing multiple victims of a perpetrator to act together, which disincentivizes retaliatory countermoves by perpetrators. In addition, combined action by victims against a perpetrator reduces disbelief by authorities as well as likelihood of reputation damage.

Callisto approaches the problem of protecting victim (and perpetrator) privacy through comprehensive use of privacy-preserving encryption technologies and user authentication practices. Personal information of users, their accounts of incidents, and the identities of perpetrators are protected before they leave the user's computer until they are decrypted on the personal workstation of the Callisto Options Counselor. In addition, even the Counselor cannot see incident or perpetrator identity information unless more than one user has identified the same perpetrator. Access to user accounts will be protected by multi-factor authentication, strong password requirements, and both e-mail and telephone contact verification.

Callisto's new offering is available to users by invitation only. Partnering organizations will provide Callisto with a list of email addresses of their members. Users with those email addresses will receive an encrypted e-mail invitation to activate accounts on Callisto's system and verify their identity. Once verified, users are free to submit incident reports, modify them, or delete them at will. Those incident reports include the identity of the accused perpetrator, which may be in one or more of several forms: a cell phone number, a social media URL, an e-mail address, and other forms that Callisto supports. When matches are identified, the Options Counselor will reach out to each victim individually (and if appropriate, may connect the group of matched victims together) to help victims of serial perpetrators find their desired pathway to justice. Information exchanged between users and the Options Counselor is protected under client privilege.

3 Callisto's Privacy Risk Management Framework

Callisto takes the responsibility of protecting user privacy seriously. Threats to privacy of information in a computer system come from a variety of adversaries: *external* adversaries that may break into a system in a variety of ways, from brute force cyber exploits to compromising user credentials; *insider* adversaries who may have authorized access to a system and use those rights inappropriately; and *impostor* adversaries who pretend to be genuine system users in order to see what can be learned.

Our early security stance assessment uses a lightweight form of the National Institute of Standards and Technology Risk Management Framework (NIST RMF) assessment [6]. We plan a full RMF assessment as product development continues, and throughout product lifetime. In addition, we continually self-test our systems for security vulnerabilities, and continuously apply patches and updates. Here, we preview our initial NIST RMF assessment by providing an overview of the information assets we aim to protect for users and perpetrators, and our approaches for protecting those assets. We describe these assets roughly in the order they appear in the user experience flow of our product.

3.1 Sensitive Information Assets Held by Callisto

Invited Users Callisto partners with different organizations to provide access to the platform and counseling services (as applicable). *Invited Users* are members who belong to these organizations and are email-invited to have access to Callisto. The information Callisto learns about Invited Users includes:

• *Invitation email address* - email addresses are sensitive: they may include names, and may be correlated with other external information held by a potential adversary

• Organization name - organization names are sensitive: the relationship between an invited user and an organization may not be public knowledge. In addition, the customer relationship with Callisto may not be public knowledge

In addition, the message we send to invite a user is potentially sensitive, so we protect

• Account outreach message - the mail we send to invite each user to register their account

Activated Users An *Activated User* is an Invited User who visits the Callisto site and registers their account. The information Callisto learns about Activated Users includes:

- Account username will be the email address provided by the sponsoring organization of the user
- *User's name* is highly sensitive for a participant in our system. An activated user might be inferred to be a victim of, or potential reporter of, an assault or harassment incident—something the user may wish to keep private
- Phone number is also highly sensitive information about a user
- Access to Callisto is sensitive, so we protect user access in multiple ways: we protect each Activated User's passphrase and use additional authentication factors to verify user access rights
- Gender is treated as sensitive because the number of users in our system of some genders may be small, making re-identification easier for an adversary

Escrowed Users Escrowed Users are Activated Users who have submitted one or more records of sexual assault or harassment. If an adversary discovers that a user is escrowed, or exfiltrates records, it can cause significant damage to the user and to the perpetrator. This information includes:

- *Incident records* that describe details of harassment or assault events, which are highly sensitive information
- Perpetrator identities, which are also highly sensitive, and should be kept private even from Callisto Options Counselors until more than one user reports the same perpetrator

Matched Users A *Matched User* is an Escrowed User who has a perpetrator match with another user. The Options Counseling process at Callisto is a workflow that supports each Matched User through a sequence of steps to reach a personalized desired outcome. The record of progress of this workflow is also sensitive:

- Transitions from one counseling step to the next are sensitive because they reveal information about how an adversary might best disrupt the Counseling process
- *Timestamps* of those step transitions are also sensitive, because they might be combined with external information to learn information about users
- Encrypted counselor notes are sensitive because they may contain personally identifiable information about users and perpetrators

Callisto Options Counselor Credentials The Options Counselor is retained by Callisto to aid Matched Users in their journey to a desired outcome. Because this counselor needs access to all details of incidents, their access credentials are quite sensitive. We protect:

- Account passwords, even though a Counselor's login alone does not allow access to sensitive user or incident information described above
- Secret access keys that the Counselor uses to access to the encrypted information of users and incidents to which that counselor is assigned

Callisto Information Technology Support Personnel Callisto IT support personnel have no access to the encrypted information about users and incidents described above. However, an adversary that compromises the credentials of an IT personnel may be able to corrupt or destroy that information, and thus harm our users and our business.

• Access to Callisto systems by IT personnel is secured in the same way as access by Users and Counselors.

Callisto Data Analysts Callisto uses statistics about incidents and progression through the options counseling workflow in order to measure user impact. For this reason, Data Analysts have access to options counseling workflow state statistics, but no access or cryptographic keys to user information, perpetrator identities, or incident records. However, even this statistical information may be sensitive, so we secure access by our analysts in the same multi-factor way that we use for users, Counselors, and IT personnel.

3.2 Technical approaches for securing sensitive information at Callisto

Table 1 below describes, for each asset type above, how we protect the asset type, as well as who has access under what conditions.

4 Cryptographic Design for Incident Records and Perpetrator Identities

4.1 Roles in our Cryptographic Design

Data submission involves interaction between several parties: the user's browser, the Callisto application server, and the Callisto key server. Data is stored within a relational database on the Callisto application server. The key server serves two roles: it stores a predetermined key whose purpose is explained below, and it authenticates Callisto users during the login process. Data is reviewed by the Callisto Options Counselor in their browser.

4.2 Cryptographic Components

The system is designed to ensure that a compromise of one of the Callisto servers, the application server or the key server, reveals no information about non-matched records. To achieve this goal the Callisto system uses the following cryptographic components:

• Shamir Secret Sharing: let s be a secret key. Shamir Secret Sharing is a technique that lets us split s into many shares s_1, s_2, \ldots, s_n so that (1) a single share reveals nothing about s, and (2) when two shares become public, anyone can reconstruct the secret s [7]. Briefly, to create shares of s we generate a random line in a plane

Information Asset Type	How Protected	Decryptable by
Invitation e-mail address	Public key cryptosystem	Options Counselor
Sponsoring organization iden-	Public key cryptosystem	Options Counselor
tity		
Account outreach message	Commercial secure messaging service	Invited User
Account username	Hashed repeatedly using SHA-2	N/A
User's personal data	Public key cryptosystem	Options Counselor
User authentication process	Multi-factor authentication service	N/A
User passphrase	Hash stretching, e.g. PBKDF2	N/A
Incident records	see Cryptographic Design, below	Options Counselor
Perpetrator identity	see Cryptographic Design, below	Options Counselor
		(only after a match)
Step metadata, notes	Public key cryptosystem	Options Counselor
Counselor secret keys	Password vault, possibly secret sharing	Options Counselor
Callisto employee passwords	Hash stretching, as for Users	N/A
Callisto employee authentica-	Multi-factor, as above	N/A
tion		

Table 1: Callisto Sensitive Information Assets

of possible secret shares whose y-intercept is the secret s. The shares of s are points on this line. A single point reveals nothing about the line, but two points reveal the line and thus enable computing its y-intercept.

- Oblivious pseudo-random functions (OPRFs). An OPRF uses a secret key k_s to map a value x to a pseudorandom value $\hat{x}[5]$. This secret key k_s is stored on the Callisto key server. A client who has an input x can interact with the Callisto key server to obtain \hat{x} . The "oblivious" property refers to the fact that in this process, the key server learns nothing about x, yet the client learns \hat{x} . We stress that this process is deterministic: evaluating the OPRF at the point x (using the key k_s) always results in the same pseudorandom value \hat{x} .
- Symmetric encryption. For a given secret key k and message m we will use c = E(k, m) to denote the encryption of m using key k. We will use D(k, c) to denote the decryption process. Callisto uses libsodium's default implementation for symmetric encryption, but will be moving towards a NIST-approved algorithm for the product.
- Public key encryption. We will use $c = \mathcal{E}(pk, m)$ to denote the encryption of m using public key pk, and $\mathcal{D}(sk, c)$ to denote the decryption of c using the corresponding secret key sk. Callisto uses libsodium for public key operations, but will be moving towards a NIST-approved algorithm for the product.

4.3 The Data Submission and Protection Process

The Callisto key server is initialized to hold the OPRF secret key k_s . The database server holds no secrets. To simplify our description here, we describe a single Callisto Options

Counselor. That Counselor generates a key pair pk and sk for a public-key encryption scheme and makes the public key pk available to the public.

With this setup complete, we informally describe the cryptographic portions of workflows for (1) record submission, (2) perpetrator matching, and (3) revealing information to the Counselor in our system at a high level. Full details, along with a cryptographic security model, will be available in a forthcoming paper.

Submitting an Incident Record. The user's computer (the client) collects the record details from the user and formats it into a serialized record structure denoted Record. This structure contains the user's identity U and the perpetrator's identity P, along with other details of the incident.

Next, the user authenticates to the key server, and once authenticated, the user's client interacts with the oblivious pseudo-random function (OPRF) system on the key server to transform the *low-entropy* perpetrator's ID P into a *pseudorandom* value \hat{P} with sufficient entropy for use in our secret sharing approach. During this step, the key server learns the identity of the user, but learns nothing about P from the user's client. Only the user's client learns \hat{P} .

The client then creates a secret share. It uses \hat{P} to derive three 256-bit pseudorandom quantities (a, k, π) using the key derivation function in libsodium. The first two quantities define a line equation Y = aX + k whose y-intercept is k. The client evaluates this line equation at the point X = U to obtain s = aU + k. The pair (U, s) is one share of a Shamir secret sharing scheme for the secret k. All arithmetic operations are performed modulo a prime number; Callisto uses the prime $p = 2^{256} + 297$.

Finally, the client encrypts Record using a fresh random record key k' to obtain an encrypted record eRecord = E(k', Record). It then encrypts k' twice, once using the key k generated above from \hat{P} , and once using a user key k_U which is discussed further below:

$$c' \leftarrow E(k, k'), \qquad c_U \leftarrow E(k_U, k').$$

All these symmetric encryptions are done using authenticated encryption with associated data (AEAD) where π is used as the associated data. The client then performs one more encryption, encrypting the triple (U, s, c') under the Options Counselor's *public key pk* to obtain a doubly-encrypted ciphertext:

$$c = \mathcal{E}(pk, (U, s, c')).$$

The client authenticates to the Callisto database server and sends it the tuple

$$(\pi, c, c_U, eRecord).$$
 (1)

The database server stores this record in its database and sends an acknowledgement to the user's browser.

On its own, this tuple (1) reveals nothing about the **Record**. Not even the Options Counselor can decrypt **eRecord**, because there is no way for them to construct the encryption key k. Moreover, if a user submits two records about the same P, this second record will result in the same share (U, s) as the first record, and thus nothing new is revealed about P. Finally, nothing about the submission process reveals anything to the user's browser about other incidents or perpetrator identities.

The user key k_U makes it possible for the user to update the record after the initial submission, if needed. The key k_U is generated on the user's client at initial submission time and the user is asked to write down this key as a sequence of four letter words. When the user needs to update the submission, the user types in this key and the user's client uses it to update the encrypted record eRecord. The system locates the record to update using π , which is derived from the perpetrator identity provided by the user using the OPRF.

Perpetrator Matching. The Callisto database server periodically performs an offline match search. If it finds two records with the same π component, it identifies these records as a match, because they share a common perpetrator. It then notifies the Options Counselor about the match. Note that matching is done without the database server having access to perpetrator identities or incident records in unencrypted form. Thus no adversary capable of penetrating the database server can learn anything about perpetrator identities or incidents from the matching process.

Revealing information to the Counselor. When the database server identifies a match, it contacts the Options Counselor who retrieves the relevant records and decrypts them using her secret key. If the records are from different incidents, the Options Counselor obtains

$$(U_1, s_1, c'_1, eRecord_1)$$
 and $(U_2, s_2, c'_2, eRecord_2)$

where $U_1 \neq U_2$. By combining the shares (U_1, s_1) and (U_2, s_2) , the Options Counselor's browser can recover the secret key k and then decrypt c'_1 and c'_2 using this key. From this it can decrypt eRecord₁ and eRecord₂. This reveals Record₁ and Record₂ in the clear, including incident details, user identities, and perpetrator identities. The Options Counselor then takes appropriate steps to contact those users and begin the resolution process.

5 Additional Details

This paper is too brief for an exhaustive description of our cryptographic approach to protecting incidents, perpetrators, and connections to the Escrowed Users reporting them. However, we include here some additional details that may be of interest to readers.

Identifying Perpetrators. Escrowed Users may identify Perpetrators using one or more diverse credentials such as social media URLs, phone numbers, or email addresses. In our system, we insist on the use of such (relatively) unambiguous identifiers. To allow for this diversity of credentials, π and s are not scalars. Instead, they are *vectors* of values $\vec{\pi}$ and \vec{s} , where each component in these vectors corresponds to a particular predetermined type of identifying credential. We say that two records $(\vec{\pi}_1, \ldots)$ and $(\vec{\pi}_2, \ldots)$ are a match if the vectors $\vec{\pi}_1$ and $\vec{\pi}_2$ match on at least one component. Moreover, the Callisto Options Counselor workstation can fill in additional components in the $\vec{\pi}$ and \vec{s} vectors for an incident once such a match is determined, because they have access to the necessary resources. Thus Callisto *propagates* perpetrator identities, using the human judgment of our Counselor, to achieve more complete identity credential vectors for perpetrators.

Privacy Roots of Trust in Callisto Every system has one or more *roots of trust*: one or more components that are assumed secure in certain ways. Briefly, our system assumes the following privacy protecting roots of trust:

- The user's browser (and computer) are one root of trust in that we assume no adversary has compromised that component with the intent of observing interactions with our system. In other words, protecting against system adversaries with vantage point on the user's computer is *out of scope* for our system. Users are responsible for adequately protecting the passphrase they use to log in, as well as the devices and accounts they use for multi-factor authentication
- Above, we described a system using a single key server. The OPRF key is a highly sensitive secret in our system. If this key is exposed to an adversary, then that

adversary can unmask records in the database by performing an exhaustive search over potential perpetrator identities. In addition, if this key is lost, then matching post-loss perpetrator identities to pre-loss identities is impossible. To further protect the OPRF key, our system uses two servers, each of which keep a single cryptographic share of the key, but not the whole key. Thus key theft requires compromise of two distinct servers with different administrators, and possibly running different operating systems. This *split server* is another root of trust of our system. One of these servers is a dedicated, highly protected physical server. The other will be a virtualized server that is hosted on a separate cloud provider. To prevent loss of the OPRF key, both servers are backed up in an encrypted backing store. To further thwart dictionary attacks, the key servers will perform rate limiting.

- The Options Counselor's workstation contains a password vault used to hold the Counselor's secret key that enables decryption of user profiles, as well as incident records and perpetrator identities (these latter two only after a match, as described above). This vault is in turn protected by a passphrase known only to the Counselor. Such passphrases are a partial root of trust for our system. We may increase security in this area by storing cryptographic shares of the Counselor's private key in a distributed fashion, and performing decryptions without bringing those key shares together in the clear, ever.
- The database server is *not* a privacy root of trust for our system, because it holds no secret keys, and because all sensitive information held there is encrypted with keys held on other components in our system.

6 Demo Application

A demo application is made available to convey our methods for client-side encryption and Shamir Secret Sharing in a user-friendly format. It can be found at

https://cryptography.projectcallisto.org.

Unlike the Callisto application, the demo is set up as a single server with the browser simulating interactions between the client and the various servers. The demo application does not reflect the full cryptographic functionality of the product. The table below indicates specific functionalities where the encryption strategies are different in the demo application versus the full product.

Functionality	Full Product	Demo
PerpID inputs	Vector of IDs	Single ID
PerpID randomization	OPRF_{k_s} with key servers	$SHA512(PerpID k_{demo})$
Symmetric Encryption	AEAD	AE
Data storage	Callisto database	Browser
Matching	Across vectors	Equality between IDs

Table 2: Implementation Differences

 k_{demo} is a pre-selected string with no correlation to k_s

7 Conclusion and Next Steps

Callisto envisions a world where sexual assault and harassment are rare and survivors are supported in their pursuit of justice. The reporting experience should be empowering for

survivors and should rebuild their sense of agency. Authorities should have the data they need to prevent assault and stop serial perpetrators.

A thoughtful cryptographic design is essential in achieving this mission. In order to create a safe space for survivors of sexual assault to come forward with their most vulnerable secret, it requires organizations like Callisto to build a technical solution that earns their trust.

As we expand to serve and empower more users, we realize that our threat models will become more sophisticated and complex. Therefore, as we expand from college campuses into industry, our solution has evolved to protect against those risks.

8 Acknowledgements

Callisto would like to thank members of its Cryptography Advisory Board, with whom we collaborated on the cryptographic design of our new product: David Archer (Galois), Dan Boneh (Stanford University), Tancrède Lepoint (SRI International), and Mayank Varia (Boston University).

We also want to acknowledge our partnership with the Software & Application Innovation Lab (SAIL) within the Hariri Institute for Computing at Boston University for their help in our software implementation of this cryptographic approach: Lucy Qin, Andrei Lapets, and Frederick Jansen.

References

- [1] EEOC Select Task Force on the Study of Harassment in the Workplace. https://www.eeoc.gov/eeoc/task_force/harassment/. [Access: March 16, 2018].
- [2] Ian Ayres and Cait Unkovic. Information escrows. *Michigan Law Review*, 111:145–196, 2012
- [3] Marcus Berzofsky Bonnie Shook-Sa Christopher Krebs, Christine Lindquist and Kimberly Peterson. Campus Climate Survey Validation Study Final Technical Report. https://www.bjs.gov/content/pub/pdf/ccsvsftr.pdf, 2016. [Access: March 26, 2018].
- [4] Susan Chibnall Reanne Townsend-Hyunshik Lee Carol Bruce David Can-Bonnie Fisher and Gail Thomas. Report on the AAUCampus Sexual Assault and Sexual Misconduct Climate Survey on //www.aau.edu/sites/default/files/AAU-Files/Key-Issues/Campus-Safety/ AAU-Campus-Climate-Survey-FINAL-10-20-17.pdf, 2017. [Access: March 27, 2018].
- [5] Katrina Ray Ryan Speers-Brian Vohaska Jonathan Burns, Daniel Moore. Ec-oprf: Oblivious pseudorandom functions using elliptic curves. IACR Cryptology ePrint Archive, 2017.
- [6] NIST. Risk Management Framework. https://csrc.nist.gov/projects/risk-management/risk-management-framework-(RMF)-Overview/. [Access: March 21, 2018].
- [7] Adi Shamir. How to share a secret. Communications of the ACM, 22(11):612–613, 1979.