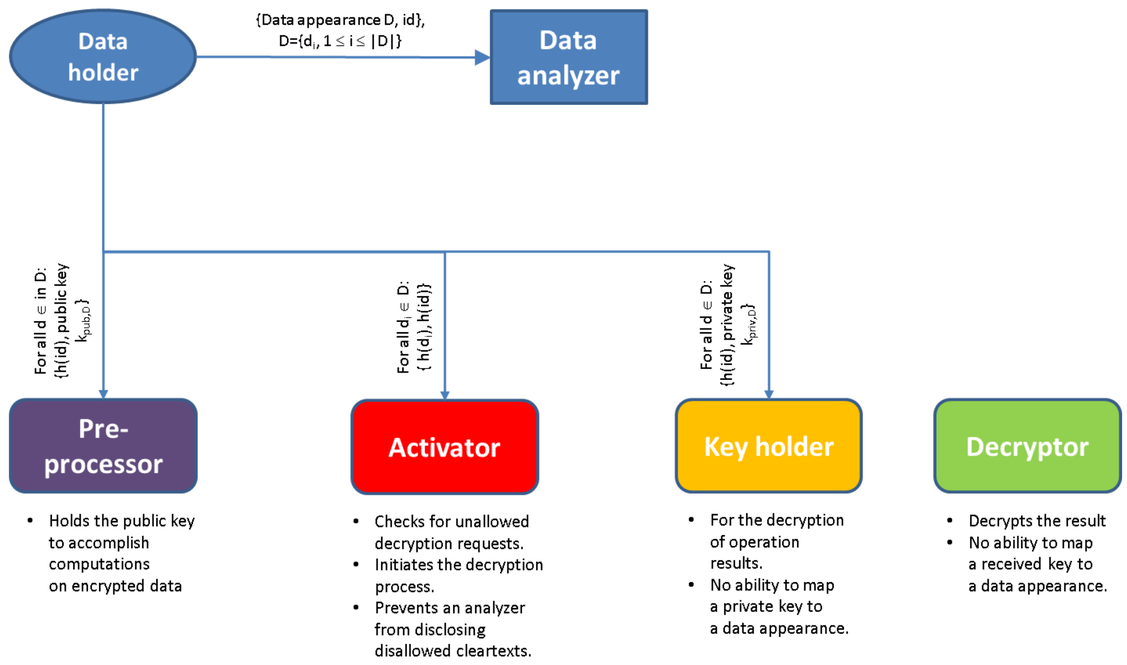
A Framework Design for Privacy-Preserving Computation on Shared Data

1. MOTIVATION

When information is shared among multiple parties for data analysis, the objective of the data holders in keeping cleartexts confidential for privacy collide with the objective of the data receiver in the highest achievable utility. Consider the following application scenario: Data holders want to share log files (e.g., see [1]) with a centralized analysis entity, called data analyzer. The data holders are interested in the results of the analysis performed by the analysis entity. The analysis entity profits from collecting log files from multiple sources because it gives him a broader view on the field, e.g. on the current situation regarding certain attacks in a business area. On the other hand, the data holders want to keep their data confidential. One reason is the privacy concerns of the data owners. This contradicts the data analyzer's interest in data utility. The described contradiction may prevent different parties from information sharing and hence, decreases the chances to profit from the analysis of data collected from multiple parties.

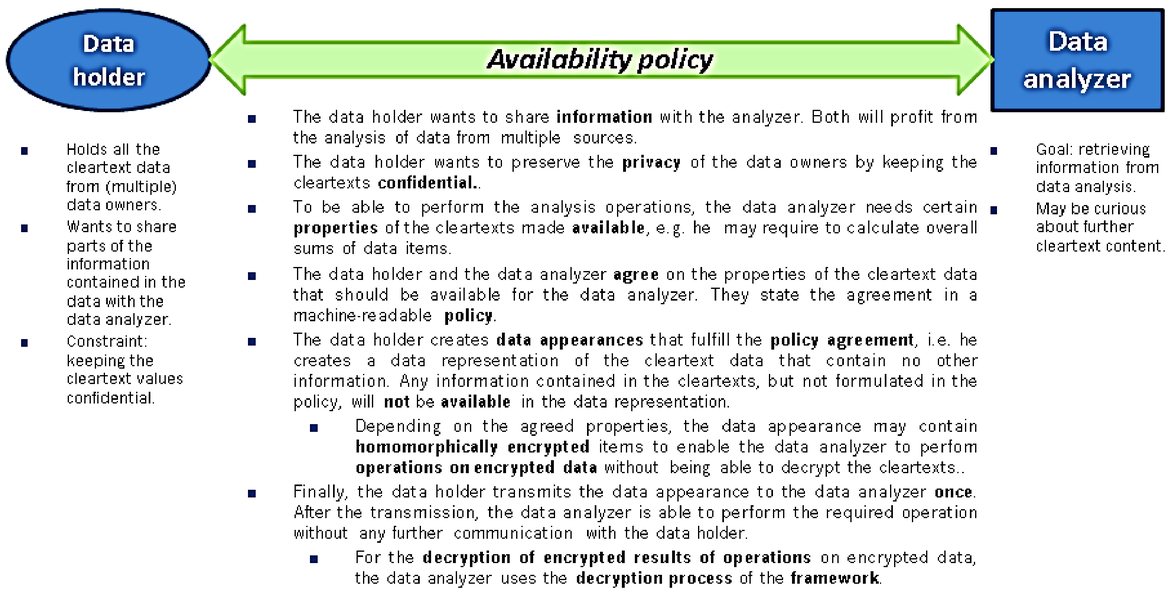
We propose a solution that meets both the privacy and utility needs of the information sharing parties. We assume that he data holders and the data analyzer agrees on the utility required for data analysis in mutual agreements called policies. A policy is a set of conditions to be fulfilled by the exchanged data in order to meet all the utility options satisfying the analyzer's and the holders' requirements. This may include the ability to calculate overall sum of numerical data for statistical analysis, or to disclose IP addresses in a log file under certain conditions, or to check encrypted data for equality. **The conditions stated in a policy must be selected very carefully to ensure that the privacy of the data owners will be preserved.** For formulating such policies, we refer to [2]. Before sharing the data with the analyzer, the data holder must transform the data to a data appearance that meets the formulated policy. Depending on the use case and the sensitivity of the data under consideration, the selection of the provided utility must keep most of the information confidential so that an attacker would not be able to use the utility properties of the data appearances for attacks utilizing information linkage or correlation with external information. In order to solve the problem described above, one approach is the utilization of homomorphic encryption. In homomorphic encryption, the cleartext data *p1; p2* are encrypted with an encryption mechanism *E* to *E(p1); E(p2),* respectively. *E* ensures that certain operations can be performed on *E(p1)* and *E(p2)* and are equivalent to operating on the plaintexts *p1* and *p2*. That means, *E(p1)*E(p2) = E(p1+p2)* for appropriate operations * and +. Note that the result of the encrypted computation is encrypted. In the above-mentioned application scenario, the data holder would use an appropriate homomorphic encryption mechanism to encrypt the cleartext content of the log files to data appearances. Depending on the selected homomorphic encryption mechanism, this would allow the analyzer to perform certain operations on the received encrypted data. To disclose the result of an operation to the analyzer, it must be decrypted. This usually requires knowledge of the secret (private) key and comes together with the fact that having access to the private key enables the data analyzer to decrypt the content of the data appearance and hence, violating the privacy of the data owners.

3. PROTOCOL INITIALIZATION AND DATA DISTRIBUTION



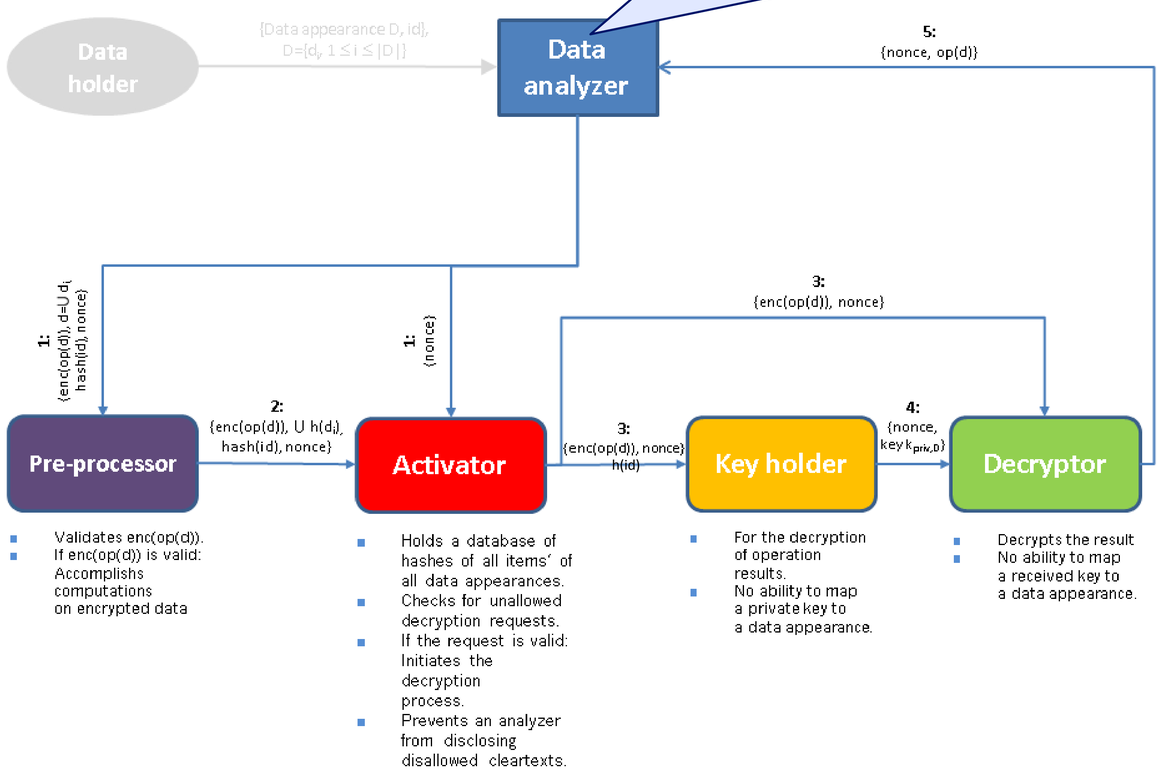
Security Assumptions:

- Honest-but curious adversarial model.
- The pre-processor has no access to private keys.
- The key holder and the decryptor cannot map an arbitrary d ∈ D to the correct private key.



4. PERFORMING OPERATIONS AND RESULT DECRYPTION

- Given: Data appearance D = {d_i, 1≤ i ≤ |D|}. d_i is the a homomorphically generated ciphertext enc(p_i)) of a plaintext p_i,.
- Goal: compute $p_k + p_l$ by computing $d_k * d_l$ for appropriate **+ and *.**
- Problem: Confidentiality of p_k, p_l.
- Solution: The data analyzer **is not allowed to decrypt,** but he requests for a decrypted result.
- Constraint: The data analyzer should not ne able to cheat using homomorphic properties of the encryption mechanism (here: Paillier [3]).



2. STAKEHOLDERS

Data holder

- Holds all the cleartext data.
- Wants to share parts of the information contained in the data with the data analyzer.
- Constraint: keeping the cleartext values confidential.

Data analyzer

- Goal: retrieving information from data analysis.
- May be curious about further cleartext content.

Pre-processor

- Holds the public key to accomplish computations on encrypted data.

Activator

- Checks for unallowed decryption requests.
- Initiates the decryption process.
- Prevents an analyzer from disclosing disallowed cleartexts.

Key holder

- For the decryption of operation results.
- No ability to map a private key to a data appearance.

Decryptor

- Decrypts the result
- No ability to map a received key to a data appearance.

5. SOURCES

[1] A. Slagell and W. Yurcik, "Sharing computer network logs for security and privacy: a motivation for new methodologies of anonymization, " Security and Privacy for Emerging Areas in Communication Networks, 2005. Workshop of the 1st International Conference on, 2005, pp. 80-89. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1588299&isnumber=33466

[2] S. Kasem-Madani and M. Meier, "Definition of Availability Policies for Data Pseudonymization Using XACML," Aug. 2015, IFIP Summer School on Privacy and Identity Management. Available: http://www.ifip-summerschool.org/

[3] P. Paillier, "Public-Key Cryptosystems Based on Composite Degree Residuosity Classes," in Advances in Cryptology, EUROCRYPT 99, ser. Lecture Notes in Computer Science, J. Stern, Ed. Springer Berlin Heidelberg, 1999, vol. 1592, pp. 223–238. Available: http://dx.doi.org/10.1007/3-540-48910-X16

Contact:

Saffija Kasem-Madani, University of Bonn
Friedrich-Ebert-Allee 144, 53113 Bonn
Germany

Tel.: 0228/73-54246

kasem@cs.uni-bonn.de
https://net.cs.uni-bonn.de/wg/itsec/staff/saffija-kasem-madani/