The Security of a Mix-Center Based on a Semantically Secure Cryptosystem

Douglas Wikström

Swedish Institute of Computer Science (SICS)
douglas@sics.se

Abstract. We introduce a definition of a re-encryption mix-center, and a definition of security for such a mix-center. Then we prove that any semantically secure public key system, which allows re-encryption, can be used to construct a secure mix-center.

1 Introduction

The notion of a mix-net was invented by Chaum [3], and further developed by a number of people. Properly constructed a mix-net enables a set of sender's to send messages anonymously. A mix-net can be viewed as an electronic analog of a tombola; messages are put into envelopes, the envelopes are mixed, and finally opened. It is impossible to tell who sent any given message. Thus the service that a mix-net provides is anonymity.

Informally the requirements on a mix-net are: correctness, privacy, robustness, availability, and efficiency. Correctness implies that the result is correct given that all mix-centers are honest. Privacy implies that if a fixed minimum number of mix-centers are honest anonymity of the sender of a message is ensured. Robustness implies that if a fixed number of mix-centers are honest, then any attempt to cheat is detected and defeated. Availability and efficiency are the general requirements on any system run on an open network.

A mix-net consists of a number of mix-centers, i.e. servers, that collectively executes a protocol. The basic idea of a mix-net, present already in Chaum's work [3], is that each mix-center receives a list of encrypted messages, transforms them, using partial decryption or random re-encryption, reorders them, and then outputs the transformed and reordered list. It should be difficult to find an element in the input list and an element in the output list that encrypts the same message. The reason for using several independent mix-centers is that it allows a sender to trust a subset of the mix-centers to ensure privacy. Later constructions have mostly dealt with robustness, availability and efficiency, which are aspects ignored by Chaum.

1.1 Previous Work and Applications of Mix-Nets

The mixing paradigm has been used to accomplish anonymity in many different scenarios. Chaum's original "anonymous channel" [3, 19] enables a sender to

A. Menezes, P. Sarkar (Eds.): INDOCRYPT 2002, LNCS 2551, pp. 368–381, 2002.

securely send mail to a receiver anonymously, and also to securely receive mail from this recipient without revealing the sender's identity. When constructing election schemes [3, 8, 21, 24, 18] the mix-net is used to ensure that the vote of a given voter can not be revealed. Also in the construction of electronic cash systems [12] mix-nets have been used to ensure anonymity. Thus a mix-net is a useful primitive in constructing cryptographic protocols.

Abe gives an efficient construction of a general mix-net [1], and argues about its properties. Jakobsson has written (partly with Juels) a number of more general papers on the topic of mixing [11, 13, 14] also focusing on efficiency, of which the first appeared at the same time as Abe's construction. Jakobsson states a result similar to our Theorem 1 given below in an informal form as Lemma 1a in [13], but for the special case where the underlying cryptosystem is El Gamal. We discuss his lemma in more detail in the next section.

Desmedt and Kurosawa [5] describes an attack on a protocol by Jakobsson [11]. Similarly Mitomo and Kurosawa [16] exhibits a weakness in another protocol by Jakobsson [13]. Pfitzmann has given some general attacks on mix-nets [23], and Michels and Horster give additional attacks in [17].

1.2 Previous Results on Mix-Centers

This work started with an attempt at writing down a formal proof of Jakobssons Lemma 1a [13] given in a slightly simplified form below. Unfortunately the statement, proof sketch, and usage of this lemma are not satisfactory.

Lemma 1a. (Jakobsson) *If the adversary can, with a non-negligible advantage ϵ over a guess uniformly at random, match any input of a mix-center to its corresponding output, then this adversarial strategy can be used as a black box to break the Decisional Diffie-Hellman assumption with a probability poly(ϵ).*

One problem is that it assumes all message variables identically and independently distributed. This model does not mirror the real world, where it is common that the adversary has some prior knowledge about the distribution of messages sent by a given party, and not all sender's should be approximated by the same distribution. Similarly it is probable that some message variables are dependent. Consider for example elections, where the votes of spouses mostly are dependent. Some problems with arbitrarily distributed messages follow.

Firstly, it is no longer clear how to state the lemma formally, since it is not clear what it should mean to "guess uniformly at random". Since the adversary knows the order of the input elements of the first mix-center he may be able to guess in different ways giving vastly different success probabilities. This is described in full detail when we argue about Definition 6.

Secondly Jakobsson assumes that the outcomes of the different copies of the message variables are all different. This allows him to say that the probability of randomly guessing a matching pair is $\frac{1}{N+2}$. This is no longer true if the number of possible messages is small. Additionally, taking this into consideration, it is not possible to argue like Jakobsson does in the argument about the $N+2$:th hybrid. He claims that if we pick new elements from the "message distribution"

the $N+2$:th hybrid will have no advantage. Consider a uniformly distributed variable over a set of only two messages. When the lists are very large it is likely that replacing all sent messages by new outcomes of the message variable, does not change the lists much, and one can not conclude that the hybrid has no advantage.

Thirdly, the proof sketch of the security of the complete mix-net of Jakobsson breaks down if we do not assume uniformly and independently distributed message variables, since he applies his lemma also to the first mix-center in the first re-encryption phase. This follows since, in the proof he permutes the input to the adversary $\mathcal{A}$ randomly, and this is not the case in the protocol, where the first mix-center in the first re-encryption phase may have partial knowledge about the distribution of the message variables.

Another problem is that Jakobsson uses Lemma 1a in his proof sketch of his Theorem 1. We discuss this issue in Section 3.3.

We conclude that a satisfactory definition, and formal proof are missing, and that some care is needed to avoid misuse of Theorem 1.

1.3 Contribution

In some cited papers above, results about security and anonymity are claimed, but a formal definition of a mix-center and a formal proof using such a definition are missing in the literature.

We provide a definition of security for a single re-encryption mix-center and show in Theorem 1 that any semantically secure re-encryption public key system can be used to construct a secure mix-center. We have restricted ourselves to mix-centers based on the random re-encryption paradigm.

We do *not* claim to give a definition of the privacy of a mix-net, since a definition of security of a complete mix-net must involve several other aspects. We highlight this in Section 3.3, where we explain two phenomena related to our theorem that occur naturally in the construction of a mix-net. One of these phenomena illustrates a misuse of Theorem 1 in the literature that to our knowledge was undetected until now.

The results we present provide some of the missing pieces in a future formal proof of security of a mix-net based on the re-encryption paradigm.

2 Notation and Definitions

We concentrate on non-uniform adversaries and denote the set of polynomial size circuit families by PC.

Let X be a random variable with probability function $p_X : \{0,1\}^n \to [0,1]$. Let M be a string describing a probabilistic circuit. We use the notation $M(X)$ for the induced random variable resulting when M is run on outcomes of X. Unless otherwise stated, all random variables are independent of all other random variables. We denote boolean values by T and F for true and false respectively.

Let Σ_N be the group of permutations on N elements. If $N(n)$ is a polynomial we sometimes write Σ_N for the family $\{\Sigma_{N(n)}\}$. We use a $\pi_N \in \Sigma_N$ both as a permutation and as a function, i.e. if l is a list $\pi_N l$ is the permuted list, and if Δ is a set of indices $\pi_N(\Delta)$ is the image of this set under the bijection defined by π_N.

We write $\mathcal{M}_n = \{0,1\}^n$, and $\mathcal{M} = \{\mathcal{M}_n\}$. We abuse notation for families of objects and the actual objects in a family, i.e. if we should write "for each n, M_n is a random variable distributed over $\mathcal{M}_n$", we instead write "M is uniformly distributed over $\mathcal{M}$". Another example of this abuse of notation is that $\Pr[D(E(m)) = m] > 1 - \frac{1}{n^c}$ in Definition 1 below should be interpreted $\Pr[D_n(E_n(m_n)) = m_n] > 1 - \frac{1}{n^c}$. The same convention is used throughout. This convention greatly simplifies the exposition.

The following definitions of a secure cryptosystem are given by Micali, Rackoff, and Sloan in [15]. The definition of semantic security given in [15] is a slightly changed version of a definition given by Goldwasser and Micali in [9]. Together these two papers give a proof of equivalence of the definition of semantic security of a cryptosystem and Definition 2 below.

Definition 1 (Public Key Cryptosystem, cf. [15]). *A* Public Key Cryptosystem *is a probabilistic Turing machine C running in expected polynomial time that on input 1^n outputs the description of two probabilistic circuits E_n and D_n of polynomial size in n such that for a polynomial $\kappa(n)$:*

1. *The encryption circuit E_n has n inputs and $\kappa(n)$ outputs.*
2. *The decryption circuit D_n has $\kappa(n)$ inputs and n outputs.*
3. *$\forall m \in \mathcal{M}$, $\forall c > 0$, $\exists n_0$ such that for $n > n_0$:*

$$\Pr[D(E(m)) = m] > 1 - \frac{1}{n^c} \ .$$

We use the notation $E(m,r)$ instead of $E(m)$ when we want to make explicit the probabilistic input r, and we assume that the number of random bits used by E is $\eta(n)$, a polynomial in n. We write $\mathcal{R}_n = \{0,1\}^{\eta(n)}$, and $\mathcal{R} = \{\mathcal{R}_n\}$.

Suppose $m = (m_1, \ldots, m_N) \in \mathcal{M}^N$, and $r = (r_1, \ldots, r_N) \in \mathcal{R}^N$. We use the notation $\overline{E}(m,r) = (E(m_1, r_1), \ldots, E(m_N, r_N))$ for element-wise encryption.

Definition 2 (GM-security, cf. [15]). *Let $(E,D) = \{(E_n, D_n)\} = \{C(1^n)\}$, where C is a public key cryptosystem, and let b be uniformly and independently distributed in $\{0,1\}$. C is* GM-secure *if $\forall m_0, m_1 \in \mathcal{M}$, $\forall T \in$ PC *and* $\forall c > 0$, $\exists n_0$ *such that* $\forall n > n_0$:*

$$\left| \Pr[T(E, m_0, m_1, E(m_b)) = m_b] - \frac{1}{2} \right| < \frac{1}{n^c} \ .$$

Another definition of security which can be proven equivalent to the other two is the following:

Definition 3 (GM-security*). *Let $(E, D) = \{(E_n, D_n)\} = \{C(1^n)\}$, where C is a public key cryptosystem, and let b be uniformly and independently distributed in $\{0,1\}$. C is* GM-secure* *if $\forall m_0, m_1 \in \mathcal{M}$, $\forall T \in$ PC and $\forall c > 0$, $\exists n_0$ such that $\forall n > n_0$:*

$$\left| \Pr[T(E, m_0, m_1, E(m_b), E(m_{1-b})) = b] - \frac{1}{2} \right| < \frac{1}{n^c} \ .$$

The following lemma is "folklore" knowledge, but for completeness we give a proof in the appendix.

Lemma 1. *Definition 3 is equivalent to Definition 2.*

3 The Security of a Mix-Center

To be able to formally prove anything about a mix-center we first define the concept of a mix-center and the right notion of security.

3.1 Definitions

The following definition captures that cryptotexts can be re-encrypted without knowledge of the private key. This property is closely related to the homomorphic property used in many papers (e.g. [10]). The by now classical El Gamal cryptosystem [7], and the recently discovered Paillier cryptosystem [20] are examples of systems that fit this definition.

Definition 4 (Re-Encryption Public Key Cryptosystem (RPKC)). *A* Re-Encryp-tion Public Key Cryptosystem *is a public key cryptosystem C that on input 1^n in addition to descriptions of E_n and D_n also outputs the description of a circuit F_n of polynomial size in n such that:*

1. *F_n has $\kappa(n)$ inputs and $\kappa(n)$ outputs.*
2. *For all $m \in \mathcal{M}$ and all $\alpha, \alpha' \in E(m, \mathcal{R})$ we have:*
   $\Pr[\alpha' = F(\alpha)] = \Pr[\alpha' = E(m)]$.

The function F above is called the "re-encryption function". As for E we use the notation $F(\alpha, r)$ instead of $F(\alpha)$ when we want to make explicit the probabilistic input r of F viewed as a deterministic circuit. Without loss of generality we can assume that E_n, D_n and F_n uses an equal number of random bits, i.e. we assume that all of the circuits use $\eta(n)$ random bits, where $\eta(n)$ is a polynomial in n. Again we use the array notation as introduced above, i.e. $\overline{F}(\alpha, r) = (F(\alpha_1, r_1), \ldots, F(\alpha_N, r_N))$, for arrays $\alpha = (\alpha_1, \ldots, \alpha_N) = \overline{E}(m)$, and $r = (r_1, \ldots, r_N) \in \mathcal{R}^N$.

Formally Definition 2 and 3 are not applicable to an RPKC. The reason is that A and T in Definition 2 and 3 respectively are given only E and not F as input. To see that this is an important detail, consider an RPKC C such that if we ignore F in the output it is GM-secure. Clearly C can encode the description

of D into the description of F, which makes C easy to break using the knowledge of F. It is however trivial to extend the definitions to be applicable also to an RPKC such that the equivalence of the definitions still holds. Thus we use the definitions as if they were defined properly for the case at hand, i.e. A and T in Definition 2 and 3 respectively take as additional input F from the output (E, D, F) of C.

Definition 5 (Re-Encryption Mix-Center (RMC)). *A* Re-Encryption Mix-Center *is a probabilistic Turing machine C_H running in expected polynomial time that on input $(1^n, N)$, N is a polynomial $N(n)$ in n, outputs descriptions of probabilistic circuits E_n, D_n, F_n, and H_n of polynomial size in n such that:*

1. *The probabilistic Turing machine $K_N(C_H)$ that, given input 1^n, simulates C_H on input $(1^n, N)$ and outputs descriptions of E_n, D_n, F_n is an RPKC.*
2. *H_n has $\kappa(n) \times N$ inputs and $\kappa(n) \times N$ outputs.*
3. *$H(\alpha) = \Pi_N \overline{F}(\alpha)$, where Π_N is uniformly distributed in Σ_N.*

We use the notation $\pi_N \overline{F}(\alpha, r) = H(\alpha)$ when we want to make explicit H's probabilistic input, i.e. π_N and r.

Note that the above is a definition of a *re-encryption* mix-center. In Chaum's [3] original construction each mix-center performed a partial decryption, and not a re-encryption. In Chaum's construction the number of input bits is not equal to the number of output bits. Also one could imagine that a mix-center received input encrypted with one cryptosystem, and produced output using another cryptosystem.

A Definition of a Secure RMC. We now introduce a notion of security for an RMC. Define a predicate ρ with regard to a given RMC taking as input a pair of lists and a pair of indices. Let $l = \overline{E}(m, r)$ and $l' = \pi_N \overline{F}(l, r')$, where $m = (m_1, \ldots, m_N) \in \mathcal{M}^N$, $r, r' \in \mathcal{R}^N$, and $\pi_N \in \Sigma_N$. Let (i, j) be a pair of indices $1 \leq i, j \leq N$. We let $\rho(l, l', i, j) = T$ if and only if it holds that $m_i = m_{\pi_N^{-1}(j)}$. The predicate is true if the encryption at index i in l and the encryption at index j in l' both encrypt the same message. It is clearly possible that there exist several pairs $(i, j_1), (i, j_2), ..., (i, j_k)$ for which $\rho(l, l', i, j_t) = T$.

The following definition says that given a secure RMC it is impossible to find a pair of indices (i, j) such that $\rho(l, l', i, j)$ holds with respect to the input l and output l' of the RMC notably better than guessing cleverly.

Definition 6 (Security of an RMC). *Let C_H be an RMC, define the family $(E, D, F, H) = \{(E_n, D_n, F_n, H_n)\} = \{C_H(1^n)\}$, and let $A \in$ PC.*

Let M be arbitrarily but independently distributed over $\mathcal{M}^N$, and let $J = \{J_n\}$, where J_n is uniformly and independently distributed over $\{1, \ldots, N(n)\}$. Define the random variables:

$$L = \overline{E}(M), \quad L' = H(L), \text{ and } \quad (I_A, J_A) = A(E, F, L, L') \ .$$

C_H is secure *if for all M and A as above $\forall c > 0$, $\exists n_0$ such that $\forall n > n_0$:*

$$|\Pr[\rho(L, L', I_A, J_A) = T] - \Pr[\rho(L, L', I_A, J) = T]| < \frac{1}{n^c} \ .$$

We argue that this is the right definition of a secure RMC as follows. Suppose that the underlying cryptosystem $K_N(C_H)$ is in some magical way perfect. That is, a cryptotext gives no information in an information theoretical sense about the encrypted message. Then an adversary clearly can not pick the second component of its output better then picking a uniformly chosen index, since the permutation Π_N, unknown to the adversary, is uniformly and independently distributed.

On the other hand the first component can still be chosen cleverly to bias the success probability. Consider for example the case where all M_i are constant, and all but one equals m_i. Then the success probability depends heavily on how the first component is chosen.

The definition states that given an adversary A that has a certain success probability, we get almost the identical success probability by using the first component of A's output and picking the second component randomly. Since we pick the second index randomly this amounts to clever guessing.

3.2 Results on the Security for an RMC

Let $K_N(C_H)$ denote the probabilistic Turing machine that given input 1^n simulates C_H on input $(1^n, N)$ to get (E_n, D_n, F_n, H_n) and outputs (E_n, D_n, F_n). We are able to prove the following theorem of which Jakobssons Lemma 1a [13] could be said to be a special case.

Theorem 1. *C_H is a secure RMC if and only if for all polynomials $N(n)$ in n, $K_N(C_H)$ is a semantically secure RPKC.*

The theorem implies that if there exists a semantically secure RPKC, then the construction given in Definition 5 gives a secure mix-center according to Definition 6. We implicitly use the generalization of Definition 2 and 3 to re-encryption public key cryptosystems, as discussed in Section 3.1. We give a proof of Theorem 1 in Appendix A.

Note that the presence of the quantification over the variable N in Theorem 1 is necessary. Without it there could exist some N for which $K_N(C_H)$ outputs trivial (E, D, F). We also need that N is polynomial in n since we otherwise would be unable to perform a hybrid argument in the proof.

3.3 Definition 6 is Not Sufficient for a Mix-Net

Our results give strong evidence for the security of many constructions of mix-nets in the literature. However they do *not* imply that the mix-nets proposed in the literature are secure, since there is not even a formal definition of security of a mix-net. Neither is Definition 6 intended to serve as a definition of security of a mix-net.

To emphasize this fact we give a generalization of an attack on mix-nets of which special cases has been described by Pfitzmann [23], and Jakobsson [11]. Jakobsson also gives a solution on how to prevent this attack. We also give an example of a situation, where our results seem to be applicable but are not.

Using Malleability to Break Anonymity. The notion of non-malleability was introduced by Dolev, Dwork and Naor [6]. Informally a cryptosystem is non-malleable if, given a cryptotext $\alpha_i = E(m_i)$ of a message m_i, it is impossible to construct $\alpha'_i = E(m'_i)$, where m'_i has some non-trivial relation to m_i.

Suppose that we have a mix-net that viewed as a single mix-center is secure by Definition 6, and that the cryptotexts given as input to the mix-net are encrypted using a malleable cryptosystem. Let $\alpha_i = E(m_i)$ be the encryption of a message m_i sent to the mix-net by Alice. We want to break the anonymity of her message. To do this we construct $\alpha'_i = E(m'_i)$ by using the malleability of the cryptosystem, where $(m_i, m'_i) \in R$, and R is some non-trivial relation. Then we send $\alpha_j = \alpha'_i$ as our message.

Thus the input to the mix-net can be written as $\alpha_1, \ldots, \alpha_N$, where α_i is the cryptotext of Alice and $\alpha_j = \alpha'_i$ is our contrived cryptotext. The output in cleartext of the mix-net has the form $\hat{m}_1, \ldots, \hat{m}_N$ where $m_i = \hat{m}_l$ for some l and $m'_i = \hat{m}_k$ for some $k \neq l$. If we apply the transformation ϕ_R on each $\hat{m}_i$, where $\phi_R(m_i) = m'_i$ and $(m_i, m'_i) \in R$, we get a list on the form: $\hat{m}'_1, \ldots, \hat{m}'_N$. Note now that $\hat{m}'_k = \hat{m}_l$. This implies that it is likely that $\hat{m}_k$ is the message sent by Alice. Depending on the relation R the probability of getting an ambiguous answer is higher or lower, and several attackers using "independent" relations increase the probability of a correct guess.

Jakobssons [11] relation is identity, and Pfitzmann [23] assumes an El Gamal cryptosystem where she uses the relation $R_x = \{(m, m^x)\}$ for some fixed x. The attack clearly fails if we use a non-malleable cryptosystem and check for identical cryptotexts, and this is what Jakobsson proposes.

The conclusion is that Definition 6 is inappropriate to define the privacy of a complete mix-net. A definition of privacy of mix-nets must allow adaptive attacks like the above, and must be defined in a multi-party setting.

Using Malleability to Break Robustness. A frequently used paradigm to achieve efficient and robust mix-net protocols is repetition. Consider the following game, where we let the underlying cryptosystem be the El Gamal system.

Let $m = (m_1, \ldots, m_N)$ be an array of cleartexts, let $\alpha = \overline{E}(m)$ be the corresponding array of cryptotexts, and let H be the output of a secure RMC. Let α' be the concatenation of h copies of the list α, and set $\alpha'' = H(\alpha')$.

Note that α'' contains a multiple of h different cryptotexts of each m_i. Suppose we are given α'' and the goal of the game is to replace all cryptotexts of any single arbitrarily chosen message m_i, with encryptions of some other message m'_i, but let the remaining set of encrypted messages be fixed. That is we must, given α'' construct a α''' such that it contains a multiple of h cryptotexts of each m_i except one m_j for which we have replaced all its cryptotexts by encryptions of some m'_j. What is the probability of success in this game?

At first it seems that if all m_i are different, then since the RMC is secure the probability should be something like the probability of guessing the position of all h copies of cryptotexts of m_i. Indeed an argument similar to this is used by Jakobsson [13] in the proof sketch of his Theorem 1.

Unfortunately this is not true in general, not even for uniformly distributed messages, as the following example shows. Suppose that the cryptosystem in use is the El Gamal system [7] over a group G. Let m_1 be uniformly and independently distributed in $G = \mathcal{M}$, let $k_1 \in G$, and $0 \neq k_2 \in \mathbb{Z}_{|G|}$ be fixed and set $m_i = k_1 m_{i-1}^{k_2}$ for all $i \neq 1$. Given an El Gamal cryptotext $\alpha_i = E(m_i)$ of m_i it is easy to compute $f(\alpha_i) = E(k_1 m_i^{k_2})$ without knowledge of the private key. Thus to succeed in our game we need only compute the list $f(\alpha)$, where we let f be defined element-wise. This maps cryptotexts of m_{i-1} onto cryptotexts of m_i except for m_N, which is mapped to an element $m'_1 \neq m_i$ for all m_i. Thus we have in effect replaced cryptotexts of m_1 with cryptotexts of m'_1 *without* identifying what cryptotexts to change.

Even though this example is not an immediate attack on any existing mix-net construction, an argument similar to this, but for a more complicated game can be found in Jakobsson's proof sketch of Theorem 1 in [13], and possibly other papers as well. We would welcome a formal proof of such claims.

4 Conclusion and Future Work

We have formalized the security of a mix-center in the re-encryption paradigm, and showed that a secure mix-center can be constructed if there exists a public key encryption system with the re-encryption property. For many mix-net constructions this is the key step in a formal proof of privacy.

A formal proof of security for a complete mix-net, in the byzantine setting is still an open question. There are many proof sketches in the literature of mix-nets, and several constructions have been broken. Since these constructions are claimed to be provably secure, we think this calls for greater attention to details. Only with formal proofs can important applications such as electronic elections be considered seriously.

An interesting future line of research is to prove a mix-net secure in the security framework of Canetti [2] or Pfitzmann and Waidner [22].

Acknowledgement

We are grateful to both Gunnar Sjödin and Johan Håstad.

References

[1] M. Abe, *Universally Verifiable mix-net with Verification Work Independent of the Number of Mix-centers*, Eurocrypt '98, pp. 437-447. 369

[2] R. Canetti, *Universally Composable Security: A New Paradigm for Cryptographic Protocols*, http://eprint.iacr.org/2000/067 and ECCC TR 01-24. Extended abstract appears in 42nd FOCS, 2001. 376

[3] D. Chaum, *Untraceable Electronic Mail, Return Addresses and Digital Pseudonyms*, Communications of the ACM, ACM 81, pp. 84-88. 368, 369, 373

[4] R. Cramer, V. Shoup, *A Practical Public Key Cryptosystem Provably Secure against Adaptive Chosen Ciphertext Attack*, Crypto '98, pp. 13-25.
[5] Y. Desmedt, K. Kurosawa, *How to break a practical MIX and design a new one*, Eurocrypt 2000, pp. 557-572. 369
[6] D. Dolev, C. Dwork, M. Naor, *Non-Malleable Cryptography*, In Proceedings of the 23rd Symposium on Theory of Computing, ACM STOC 1991. 375
[7] T. ElGamal, *A Public Key Cryptosystem and a Signiture Scheme Based on Discrete Logarithms*, IEEE Transactions on Information Theory 31, 1985, pp. 469-472. 372, 376
[8] A. Fujioka, T. Okamoto and K. Ohta, *A practical secret voting scheme for large scale elections*, Auscrypt '92, pp. 244-251. 369
[9] S. Goldwasser, S. Micali, *Probabilistic Encryption*, Journal of Computer Science 28, pp. 270-299, 1984. 371
[10] M. Hirt, K. Sako, *Efficient Reciept-Free Voting Based on Homomorphic Encryption*, Eurocrypt 2000, pp. 539-556. 372
[11] M. Jakobsson, *A Practical Mix*, Eurocrypt '98, pp. 448-461. 369, 374, 375
[12] M. Jakobsson, D. M'Raihi, *Mix-based Electronic Payments*, SAC '98, pp. 157-173. 369
[13] M. Jakobsson, *Flash Mixing*, PODC'99, pp. 83-89. 369, 374, 375, 376
[14] M. Jakobsson, A. Juels, *Millimix: Mixing in small batches*, DIMACS Techical report 99-33, June 1999. 369
[15] S. Micali, C. Rackoff, B. Sloan, *The Notion of Security for Probabilistic Cryptosystems*, SIAM J. Computing 1988, pp. 412-426. 371
[16] M. Mitomo, K. Kurosawa, *Attack for Flash MIX*, Asiacrypt 2000, pp. 192-204. 369
[17] M. Michels, P. Horster, *Some remarks on a reciept-free and universally verifiable Mix-type voting scheme*, Asiacrypt '96, pp. 125-132. 369
[18] V. Niemi, A. Renvall, *Efficient voting with no selling of votes*, Asiacrypt'94, pp. 105-116. 369
[19] W. Ogata, K. Kurosawa, K. Sako, K. Takatani, *Fault Tolerant Anonymous Channel*, ICICS '97, pp. 440-444. 368
[20] P. Paillier, *Public-Key Cryptosystems Based on Composite Degree Residuosity Classes*, Eurocrypt '99, pp. 223-238. 372
[21] C. Park, K. Itoh, K. Kurosawa, *Efficient Anonymous Channel and All/Nothing Election Scheme*, Eurocrypt '93, pp. 248-259. 369
[22] B. Pfitzmann, M. Waidner, *Composition and Integrity Preservation of Secure Reactive Systems*, 7th Conference on Computer and Communications Security of the ACM, pp. 245-254, 2000. 376
[23] B. Pfitzmann, *Breaking an Efficient Anonymous Channel*, Eurocrypt '94, pp. 332-340. 369, 374, 375
[24] K. Sako, J. Killian, *Reciept-free Mix-Type Voting Scheme*, Eurocrypt '95, pp. 393-403. 369

A Proofs

Before we prove Theorem 1 we prove some lemmas. Denote by π^0 the identity permutation and $\pi^1 = \pi$ for any permutation π. Consider the following generalization of the GM-security*, i.e. Definition 3. Throughout this section we assume that $(E, D) = \{(E_n, D_n)\} = \{C(1^n)\}$, when we write E or D.

Definition 7 (Generalized GM-security* (gGM)). *Let C be a public key cryptosystem, let $(E, D) = \{(E_n, D_n)\} = \{C(1^n)\}$, and let b be uniformly distributed in $\{0,1\}$. C is* gGM-secure *if for all polynomials N in n, $\forall \pi_N \in \Sigma_N$, $\forall m \in \mathcal{M}^N$, $\forall T \in \mathrm{PC}$, $\forall c > 0$, $\exists n_0$ such that $\forall n > n_0$:*

$$\left| \Pr[T(E, m, \pi_N^b \overline{E}(m)) = b] - \frac{1}{2} \right| < \frac{1}{n^c} \ .$$

Lemma 2. *A public key cryptosystem C is GM-secure iff it is gGM-secure.*

Proof. We see that gGM-security immediately implies GM-security*, since we may take the polynomial $N(n) = 2$ in the definition of gGM-security.

To prove the opposite direction of the lemma, we assume it is false. Then there exists a GM-secure cryptosystem C, and a polynomial N, $\exists m \in \mathcal{M}$, $\exists T \in \mathrm{PC}$, $\exists \pi_N \in \Sigma_N$, $\exists c > 0$, and an infinite set $\mathcal{N}$ such that for $n \in \mathcal{N}$:

$$\left| \Pr[T_n(E_n, m_n, \pi_{N,n}^b \overline{E_n}(m_n)) = b] - \frac{1}{2} \right| \geq \frac{1}{n^c} \ .$$

We now define an $A = \{A_n\} \in \mathrm{PC}$ that breaks the GM-security* of C. Consider a fixed $n \in \mathcal{N}$. Note that for any permutation, in particular for $\pi_{N,n}$, there exists a chain of permutations $\mathrm{id} = \pi^{(1)}, \pi^{(2)}, \ldots, \pi^{(N)} = \pi_{N,n}$, such that $\pi^{(i+1)}$ and $\pi^{(i)}$ differ only by a transposition. We get the following hybrid argument:

$$\zeta_i = \Pr[T_n(E_n, m_n, (\pi^{(i)})^b \overline{E_n}(m_n)) = b], \quad \frac{1}{n^c} \leq |\zeta_N - \zeta_1| \leq \sum_{i=1}^{N-1} |\zeta_{i+1} - \zeta_i| \ .$$

where $\zeta_1 = \frac{1}{2}$ since $(\pi^{(1)})^b = \mathrm{id}$. This implies $|\zeta_{t+1} - \zeta_t| \geq \frac{1}{Nn^c}$ for some $1 \leq t < N$. Let k_0 and k_1 be the two indices such that $\pi^{(t)}(k_0) = \pi^{(t+1)}(k_1)$ and $\pi^{(t+1)}(k_0) = \pi^{(t)}(k_1)$. Let $(E_n, m_{n,k_0}, m_{n,k_1}, \alpha_0, \alpha_1)$ be the input to A_n, where b is randomly chosen, and $(\alpha_0, \alpha_1) = (E_n(m_{n,k_b}), E_n(m_{n,k_{1-b}}))$. The circuit A_n:

1. Computes $\alpha = \pi^{(t)} \overline{E_n}(m_n)$.
2. Replaces the elements of α at positions $\pi^{(t)}(k_0)$ and $\pi^{(t)}(k_1)$ by the elements α_0 and α_1 respectively. Let the resulting vector be α'.
3. Runs $b = T_n(E_n, m_n, \alpha')$, and returns b.

It follows that the GM-security* of C is broken. □

Corollary 1. *If C is gGM-secure then, $\forall j = \{j_n\}$, where $j_n \in \{1, \ldots, N(n)\}$, $\forall \pi_N, \psi_N \in \Sigma_N$, $\forall m \in \mathcal{M}^N$, $\forall T \in \mathrm{PC}$, $\forall c > 0$, $\exists n_0$ such that $\forall n > n_0$:*

$$|\Pr[T(E, m, \pi_N \overline{E}(m)) = j] - \Pr[T(E, m, \psi_N \overline{E}(m)) = j]| < \frac{1}{n^c} \ .$$

Proof. Assume the contrary. Then there exists j, π_N, ψ_N, and $c > 0$ such that for $n \in \mathcal{N}$ the inequality above does not hold. Consider a fixed $n \in \mathcal{N}$. If we set $\zeta_{\pi_N} = \Pr[T_n(E_n, m_n, \pi_{N,n} \overline{E_n}(m_n)) = j_n]$, we have $|\zeta_{\pi_N} - \zeta_{\psi_N}| \geq \frac{1}{n^c}$.

We now construct a $B = \{B_n\} \in \mathrm{PC}$ that breaks the gGM-security of C. B_n takes input (E_n, m_n, α), where $\alpha = (\pi_{N,n}^{-1}\psi_{N,n})^b\overline{E_n}(m_n)$, and $b \in \{0,1\}$ is randomly chosen.

The circuit B_n does the following $k(n)$ times, where $k(n)$ is a polynomial: It computes $\alpha' = \pi_{N,n}\overline{F_n}(\alpha)$, and runs $T_n(E_n, m_n, \alpha')$. Then it computes the fraction s of times T_n returned j_n. If $|s - \zeta_{\pi_N}| < |s - \zeta_{\psi_N}|$ then it returns 0 and otherwise it returns 1.

Let Z_i be an indicator variable of the event that T_n outputs j_n in the i:th run. Then s is an outcome of the random variable $Z(k) = \frac{1}{k(n)}\sum_{i=1}^{k(n)} Z_i$. Thus: $\mathrm{Var}[Z(k)|b=0] = \frac{1}{k}\mathrm{Var}[Z_1|b=0] = \frac{1}{k}\zeta_{\pi_N}(1-\zeta_{\pi_N})$, and similarly for ψ_N. Let $d = \frac{1}{2}|\zeta_{\pi_N} - \zeta_{\psi_N}|$, then we have from Chebychev's bound that:

$$\Pr[|(Z(k)|b=0) - \zeta_{\pi_N}| > d|b=0] \leq \frac{\mathrm{Var}[Z(k)|b=0]}{d^2} \leq \frac{n^{2c}}{k(n)}$$

and similarly for $b=1$. Thus if we set $k(n) = 2n^{2c}$ we have:

$$\Pr[B_n(E_n, m_n, (\pi_{N,n}^{-1}\psi_{N,n})^b\overline{E_n}(m_n)) = b] \geq \frac{1}{2}$$

which breaks the gGM-security of C. □

Lemma 3. *For $m_n \in \mathcal{M}_n^N$, denote by $\Delta_i(m_n)$ the set $\{j|m_{n,j} = m_{n,i}\}$, and let Π_N be uniformly distributed in Σ_N. If C is GM-secure then $\forall i = \{i_n\}$, where $i_n \in \{1,\ldots,N(n)\}$, $\forall m \in \mathcal{M}^N$, $\forall T \in \mathrm{PC}$, $\forall c > 0$, $\exists n_0$ such that $\forall n > n_0$:*

$$\left|\Pr[T(E, m, \Pi_N\overline{E}(m)) \in \Pi_N(\Delta_i(m))] - \frac{|\Delta_i(m)|}{N}\right| < \frac{1}{n^c} \ .$$

Proof. Let $\epsilon_{\pi_N,j} = \Pr[J_T = j|\Pi_N = \pi_N] - \Pr[J_T = j|\Pi_N = \mathrm{id}]$, where we let $J_T = T(E, m, \Pi_N\overline{E}(m))$. We write Δ_i for $\Delta_i(m)$, and have:

$$\begin{aligned}
\Pr[J_T \in \Pi_N(\Delta_i)] &= \sum_{\pi_N \in \Sigma_N} \frac{1}{N!}\Pr[J_T \in \Pi_N(\Delta_i)|\Pi_N = \pi_N] \\
&= \sum_{j=1}^{N}\sum_{\pi_N \in \Sigma_N} \frac{1}{N!}\Pr[J_T = j|\Pi_N = \pi_N]\Pr[J_T \in \Pi_N(\Delta_i)|\Pi_N = \pi_N, J_T = j] \\
&= \sum_{j=1}^{N}\Pr[J_T = j|\Pi_N = \mathrm{id}]\sum_{\pi_N \in \Sigma_N} \frac{1}{N!}\Pr[J_T \in \Pi_N(\Delta_i)|\Pi_N = \pi_N, J_T = j] \\
&\quad + \sum_{j=1}^{N}\sum_{\pi_N \in \Sigma_N} \frac{1}{N!}\epsilon_{\pi_N,j}\Pr[J_T \in \Pi_N(\Delta_i)|\Pi_N = \pi_N, J_T = j] \\
&= \frac{|\Delta_i|}{N} + \sum_{j=1}^{N}\sum_{\pi_N \in \Sigma_N} \frac{1}{N!}\epsilon_{\pi_N,j}\Pr[J_T \in \Pi_N(\Delta_i)|\Pi_N = \pi_N, J_T = j]
\end{aligned}$$

since $\sum_{\pi_N \in \Sigma_N} \frac{1}{N!} \Pr[J_T \in \Pi_N(\Delta_i) | \Pi_N = \pi_N, J_T = j] = \frac{|\Delta_i|}{N}$. Thus we have:

$$\left| \Pr[J_T \in \Pi_N(\Delta_i)] - \frac{|\Delta_i|}{N} \right| \leq \sum_{j=1}^{N} \sum_{\pi_N \in \Sigma_N} \frac{1}{N!} |\epsilon_{\pi_N,j}| \leq N \max_{\pi_N,j} \{|\epsilon_{\pi_N,j}|\}$$

which is negligible since $N(n)$ is polynomial and $\max_{\pi_N,j}\{|\epsilon_{\pi_N,j}|\}$ by Corollary 1 is negligible. □

We are now ready to give the proof of Theorem 1.

Proof (of Theorem 1). First the easy direction of the proof. Suppose that C_H is a secure RMC, but $C = K_N(C_H)$ is not a GM-secure RPKC for some polynomial N. Then $\exists m_0, m_1 \in \mathcal{M}$, $\exists T \in \text{PC}$, $\exists c > 0$ and an infinite index set $\mathcal{N}$ such that for $n \in \mathcal{N}$:

$$\left| \Pr[T(E, m_0, m_1, E(m_b), E(m_{1-b})) = b] - \frac{1}{2} \right| \geq \frac{1}{n^c} \ .$$

The family $A = \{A_n\}$, where A_n given input $(E_n, F_n, \overline{E_n}(m_{n,0}, m_{n,1}), (\alpha_0, \alpha_1))$ returns the pair $(0, T_n(E_n, m_{n,0}, m_{n,1}, \alpha_0, \alpha_1))$ shows that C_H is not secure.

To prove the other direction, we assume that $K_N(C_H)$ is semantically secure for all polynomials N, but C_H is not secure. Then, using the notation of Definition 6, there exists an $A \in \text{PC}$, an infinite index set $\mathcal{N}$, and a $c > 0$ such that for $n \in \mathcal{N}$:

$$|\Pr[\rho(L_n, L'_n, I_{A_n}, J_{A_n}) = T] - \Pr[\rho(L_n, L'_n, I_{A_n}, J_n) = T]| \geq \frac{1}{n^c} \ .$$

We abuse notation and write $\rho(I, J)$ instead of the correct $\rho(L_n, L'_n, I, J)$. A probabilistic argument gives that there exists a fixed $m \in \mathcal{M}^N$ such that for $n \in \mathcal{N}$: $|\Pr[\rho(I_{A_n}, J_{A_n}) = T | M_n = m_n] - \Pr[\rho(I_{A_n}, J_n) = T | M_n = m_n]| \geq \frac{1}{n^c}$.

We define: $\zeta_{A,i} = \Pr[\rho(I_{A_n}, J_{A_n}) = T | M_n = m_n, I_{A_n} = i]$ and similarly $\zeta_i = \Pr[\rho(I_{A_n}, J_n) = T | M_n = m_n, I_{A_n} = i]$ to simplify notation in the following.

For some $1 \leq t \leq N(n)$ we have:

$$\begin{aligned} \frac{1}{n^c} &\leq |\Pr[\rho(I_{A_n}, J_{A_n}) = T | M_n = m_n] - \Pr[\rho(I_{A_n}, J_n) = T | M_n = m_n]| \\ &= \left| \sum_{i=1}^{N(n)} p_{I_{A_n}}(i)(\zeta_{A,i} - \zeta_i) \right| \leq N(n) p_{I_{A_n}}(t) |\zeta_{A,t} - \zeta_t| \ . \end{aligned}$$

We construct a $T \in \text{PC}$ that contradicts Lemma 3. The circuit T_n gets input (E_n, F_n, m_n, l'_n), where l'_n is an outcome of L'_n, computes $l_n = \overline{E}(m_n)$, runs $(i, j) = A_n(E, F, l_n, l'_n)$, and if $i = t$ it returns j, and otherwise it returns the outcome of a random variable J_n, which is uniformly and independently distributed over $\{1, \ldots, N(n)\}$.

Using the notation of Lemma 3 we have $\zeta_t = \frac{|\Delta_t|}{N}$ which gives:

$$\begin{aligned}\Pr[J_{T_n} \in \Pi_{N(n)}(\Delta_t)] &= \sum_{i \neq t} p_{I_{A_n}}(i)\frac{|\Delta_t|}{N} + p_{I_{A_n}}(t)\zeta_{A,t} \\ &= \frac{|\Delta_t|}{N} + p_{I_{A_n}}(t)(\zeta_{A,t} - \zeta_t) \ .\end{aligned}$$

Thus $|\Pr[J_{T_n} \in \Pi_{N(n)}(\Delta_t)] - \frac{|\Delta_t|}{N}| = p_{I_{A_n}}(t)|\zeta_{A,t} - \zeta_t| \geq \frac{1}{N(n)n^c}$, which contradicts Lemma 3. □

In the proof above we implicitly use an extended version of Definition 7 that is applicable to an RPKC, and use that Lemma 2, Corollary 1, and Lemma 3 hold correspondingly (see Section 3.1).

For completeness we give a proof of Lemma 1.

Proof (of Lemma 1). Suppose that a PKC C is not secure according to Definition 2, and let $T = \{T_n\}$ be the family of circuits that shows this. Then the family of circuits $T' = \{T'_n\}$, where T'_n simulates T_n on the first component of its input and returns b if T_n returns m_b is clearly not secure according to Definition 3.

For the other direction, suppose that a PKC C is not secure according to Definition 3. Then $\exists m_0, m_1 \in \mathcal{M}$, $\exists T' \in \mathrm{PC}$, and an infinite index set $\mathcal{N}$, such that for each $n \in \mathcal{N}$:

$$\begin{aligned}p_{bd} &= \Pr[T'(E, m_0, m_1, E(m_b), E(m_d)) = 1] \\ \frac{1}{n^c} &\leq |p_{01} - p_{10}| = |p_{01} - p_{11} + p_{11} - p_{10}| \leq 2|p_{t,1-t} - p_{11}|\end{aligned}$$

for some $t = \{t_n\}$, where $t_n \in \{0, 1\}$. Set $\gamma_t = \alpha$ and $\gamma_{1-t} = E(m_1)$. Then T runs $b = T'(E, m_0, m_1, \gamma_0, \gamma_1)$ and returns m_b. It follows that T breaks the security according to Definition 2. □