Privacy Amplification from Non-malleable Codes

Eshan Chattopadhyay * Bhavana Kanukurthi ** Sai Lakshmi Bhavana Obbattu *** Sruthi Sekar†

Abstract. In this paper, we connect two interesting problems in the domain of Information-Theoretic Cryptography: "Non-malleable Codes" and "Privacy Amplification". Non-malleable codes allow for encoding a message in such a manner that any "legal" tampering will either leave the message in the underlying tampered codeword unchanged or unrelated to the original message. In the setting of Privacy Amplification, we have two users that share a weak secret w guaranteed to have some entropy. The goal is to use this secret to agree on a fully hidden, uniformly distributed, key K, while communicating on a public channel fully controlled by an adversary.

While lot of connections have been known from other gadgets to NMCs, this is the first result to show an application of NMCs to any information-theoretic primitive (other than tamper resilient circuits). Specifically, we give a general transformation that takes any augmented non-malleable code and builds a privacy amplification protocol. This leads to the following results:

(a) Assuming the existence of constant rate, optimal error[1], two-state augmented non-malleable code there exists a 8-round privacy amplification protocol with optimal entropy loss and min-entropy requirement $\Omega(\log(n) + \kappa)$ (where κ is the security parameter). In fact, "non-malleable randomness encoders" suffice.

(b) Instantiating our construction with the current best known augmented non-malleable code for 2-split-state family [Li17], we get a 8-round privacy amplification protocol with entropy loss $\mathcal{O}(\log(n) + \kappa \log(\kappa))$ and min-entropy requirement $\Omega(\log(n) + \kappa \log(\kappa))$.

1 Introduction

The field of Information-theoretic Cryptography has seen a flurry of exciting research activity in recent times, specifically on the problems of Non-malleable

---

* Cornell University and IAS, Email: eshan.c@gmail.com. Research supported by NSF grant CCF-1412958 and the Simons foundation

** Department Of Computer Science and Automation, Indian Institute Of Science, Email: bhavana@iisc.ac.in. Research supported in part by Department of Science and Technology Inspire Faculty Award.

*** Department Of Computer Science and Automation, Indian Institute Of Science, Email: oslbhavana@gmail.com

† Department Of Mathematics, Indian Institute Of Science, Email: sruthi.sekar1@gmail.com.

[1] We say an ϵ-(augmented) NMC has optimal error if $\epsilon = 2^{-\mathcal{O}(message\ length)}$

Codes and Privacy Amplification. Non-malleable codes were introduced in the work of Dziembowski, Pietrzak and Wichs [DPW10] and provide an encoding mechanism with the following guarantee: errors caused to the codeword will render the underlying data either independent of the original encoded message or leave it unchanged. NMCs are defined with respect to a class of tampering families $\mathcal{F}$. The class of tampering families most relevant to this work is the "2-Split-state" family where the codeword consists of two states L and R and the tampering family consists of two functions f and g, each acting independently on L and R respectively. A parameter of importance for any non-malleable coding scheme is its *rate* ($= \frac{\mathsf{message\ length}}{\mathsf{codeword\ length}}$). Of late, there has been tremendous research in building non-malleable codes with low-rate for various tampering function families, in particular, the 2-Split-state model. Researchers have also explored connections of other primitives, such as "2-source Non-malleable Extractors" to NMCs. Yet, in spite of the interest in non-malleable codes, to this date, to the best of our knowledge, there hasn't been a single application of non-malleable codes to building any other information-theoretic primitive. The main challenge of using NMCs is that they are typically secure only with respect to a restricted class of tampering functions (such as 2-split state tampering). Most natural applications will require arbitrary tampering of the entire codeword. We overcome this challenge and present an important application of NMCs, namely to building Privacy Amplification Protocols.

We now describe the problem of Privacy Amplification, introduced by Bennett, Brassard and Robert [BBR88]. In this setting, we have two parties, Alice and Bob, who share a common string w, that is only guaranteed to be entropic. The main question that is asked is the following: How can Alice and Bob use w to communicate over a public channel that is fully controlled by a computationally-unbounded adversary, Eve, and still agree on a key K whose distribution is close-to-uniform? This problem has received renewed attention in recent years. While building privacy amplification protocols, there are two main objectives that researchers have tried to meet: a) build protocols with as low a round complexity as possible and b) extract a key K that is as long as possible. To achieve the latter objective, a natural goal is therefore to minimize the *"entropy loss"* that occurs due to the protocol.

Our main result in this work is that we show how to build privacy amplification protocols from non-malleable codes, specifically those with the so-called "augmented" security which we explain later. The protocol has 8 rounds and its entropy loss of is related to the rate of the non-malleable code. Furthermore, even though our main protocol is presented in terms of non-malleable codes we can also use the weaker notion of Non-malleable Randomness Encoders in the place of non-malleable codes. Non-mallebale Randomness Encoders (NMREs) were introduced by Kanukurthi, Obbattu and Sekar [KOS18] and, informally, allow for non-malleable encoding of "pure randomness". There is evidence to suggest that it is easier to build NMREs (with good parameters) than NMCs: specifically, while we know how to build constant-rate NMREs in the 2-Split

State Model, a similar result for NMCs has proven ellusive in spite of significant interest and effort in the research community.

1.1 Prior Work on Privacy Amplification (PA)

Recall that the goal of privacy amplification is to enable two parties with a weak (entropic) secret w to agree on a random key K whose distribution is close to uniform. The protocol communication takes place in the presence of a computationally unbounded adversary, Eve, who has complete power to insert, delete or modify messages. Intuitively, a privacy amplification protocol is considered to be secure if any such adversarial tampering of the communication is either detected by one of the honest parties or, if undetected, both parties do agree on the same "secure" key, i.e., one that is guaranteed to be close to uniform from the Eve's point of view. It is no surprise that strong randomness extractors (introduced by Nissan and Zuckerman [NZ96]), which transform non-uniform randomness into uniform randomness by using a short uniformly chosen *seed*, play a huge role in the design of privacy amplification protocols. Specifically, in the setting where Eve is a passive adversary [Mau93,BBR88,BBCM95], strong randomness extractors offer a one round solution to the above problem, which is optimal (in terms of entropy loss and min-entropy requirements).

In the setting where Eve is an active adversary, a one-round solution to the problem was first given by Maurer and Wolf [MW97] with min-entropy requirement of $k_{min} > 2n/3$, where k_{min} is the starting min-entropy requirement and n is the length of w. This was later improved in Dodis, Katz, Reyzin and Smith [DKRS06] (with min-entropy requirement of $k_{min} > n/2$). The negative results by [DS02,DW09] show that there is no non-interactive (one-round) solution for this problem when the entropy of the weak secret is $k_{min} \leq n/2$. Hence, for $k_{min} \leq n/2$, researchers explored the use of interaction to design privacy amplification protocols.

In the interactive setting with an active adversary, there are two major lines of work. The first line of constructions began with the protocol given by Renner and Wolf [RW03] who gave a protocol with an entropy loss of $\Theta(\kappa^2)$ and takes $\Theta(\kappa)$ rounds of communication, where κ is the security parameters. This was generalized by Kanukurthi and Reyzin [KR09]. In [CKOR10], Chandran, Kanukurthi, Ostrovsky and Reyzin, used codes for Edit-distance Metric with optimal rate to achieve the first protocol with an entropy loss of $\Theta(\kappa)$. The high-level approach of Renner and Wolf's protocol, which was followed in subsequent works, was to first build an "interactive authentication protocol" which authenticates the message bit-by-bit. This authentication protocol is then used to authenticate a seed to a randomness extractor which is then used to extract the final key K, thereby achieving privacy amplification. A natural limitation of this approach is that it is highly interactive and requires $\Theta(\kappa)$ rounds.

The second line of constructions began with the privacy amplification protocol given by Dodis and Wichs [DW09]. They give an efficient two-round construction (i.e., with optimal round complexity) which has an entropy loss of $\Theta(\kappa^2)$. This work also introduces "*seeded Non-malleable extractors (NME)*", which has the property that the output of the extractor looks uniform, even given its value on a related seed. Their approach for building two-round privacy amplification protocols roughly works as follows: first, they send a seed to a NME which is used to extract the key (k) to a non-interactive one-time message authentication code. k is then used to authenticate a seed s to an extractor. The final shared key K is evaluated by both parties, unless any tampering is detected, to be $\mathsf{Ext}(w; s)$. In short, the approach of Dodis and Wichs leads to a Privacy amplification protocol with optimal round complexity of 2. Further, [DW09] give an existential result that if one can efficiently construct non-malleable extractors with optimal parameters, we get a two-round privacy amplification protocol with entropy loss $\Theta(\kappa)$ and min-entropy requirement $\mathcal{O}(\kappa + \log n)$. Subsequent to the existential construction of Privacy Amplification given in [DW09], there was focus on improving the parameters by giving explicit constructions of seeded non-malleable extractors [DLWZ11,CRS11,Li12a,Li12b,Li15,CGL16,CL16,Coh16,Li17]. While all these constructions give a 2-round privacy amplification protocol with optimal entropy loss, the min-entropy requirement is not optimal.

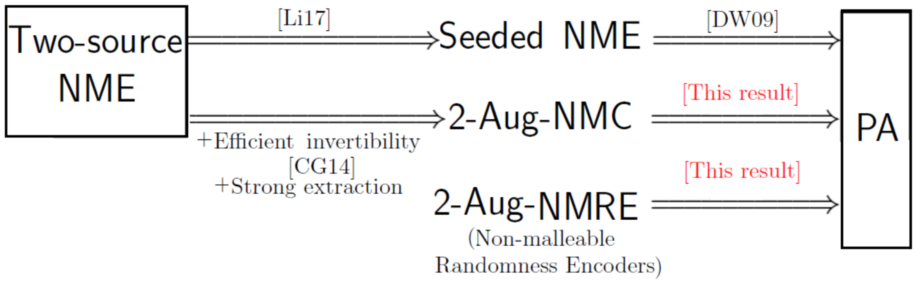
1.2 Overview of Research on NMCs

We now give a brief overview of Non-malleable Codes. NMCs, introduced by Dziembowski, Peitrzak and Wichs, guarantee that a tampered codeword will decode to one of the following:

- $\perp$ i.e., the decoder detects tampering.
- the original message m itself i.e., the tampering did not change the message
- something independent of m

Since, as observed in [DPW10], NMCs cannot be built to be secure against arbitrary, unrestricted tampering, researchers have explored the problem of building NMCs for various classes of tampering families $\mathcal{F}$. The most well-studied model is the "$t-$split state" model where a codeword consists of t states $(C_1, \ldots C_t)$ and the tampering functions consists of t functions $f_1, \ldots, f_t$. The model permits independent tampering of each C_i via the function f_i. (Each f_i itself is not restricted in any way and, therefore, the model enables arbitrary but independent tampering of each state.) Over a series of works researchers have built NMCs for varying values of t, where $t = 2$ represents the least restrictive model of tampering and $t = n$, for codeword length n, represents the most restrictive model [DPW10,CG14,ADL14,CZ14,ADKO15,AGM$^+$15,Li17,KOS17,KOS18]. At the same time, researchers have also focused on building constructions with good (low) rate. To this date, the problem of building constant rate non-malleable codes in the 2-split state model remains open. In [KOS18], the authors introduced a notion called "Non-malleable Randomnes Encoders" which allow for non-malleably encoding "pure randomness". Furthermore, they also present a construction of an NMRE with a constant rate of $\frac{1}{2}$. As we will explain later, the rate of our NMCs/NMREs is closely linked to the entropy loss of the resulting privacy amplification protocol.

Researchers have also explored connections of NMCs to other primitives, as demonstrated by the following picture.



However, somewhat surprisingy, to the best of our knowledge, there isn't a single application of Non-malleable Codes to *any* information-theoretic primitive. One of the reasons for this is that the split-state model doesn't allow for arbitrary tampering when the whole codeword is visible, which most natural applications might require. In this work, we present an application of augmented NMCs (and NMREs) to Privacy Amplification. (Augmented non-malleable codes are secure even if one of the states is leaked to the adversary after the tampering.) We now give an overview of our techniques to build privacy amplification.

1.3 Technique for building PA from NMC.

In this work, we deviate from the approaches due to Renner and Wolf (of bit-wise authentication) as well as Dodis and Wichs (of using Non-malleable Extractors) and present a new technique to obtain privacy amplification from (augmented) Non-malleable Codes. (We will use certain elements of Renner and Wolf's approach, which we will describe shortly.) Just as in prior works, the heart of the protocol consists of an authentication protocol from which we can easily obtain a privacy amplification protocol. So for the rest of this discussion, we restrict our attention to interactive authentication and describe our protocol for the same at a high level. Suppose Bob wants to authentically send a message m to Alice. Alice intiates the protocol by picking a random key k for the MAC, encodes it into (L, R) using a non-malleable code and sends it to Bob. Bob can then authenticate his message using the received key for the MAC and send the message and the tag to Alice. In order to be able to use the MAC security, we must ensure that the MAC key k looks uniform even given the information leaked through the communication channel. It seems natural that the use of non-malleable codes would ensure that even if Eve tampers the channel, Bob would either get the original key or an independent key k. In such a case, the tag evaluated using the MAC key k' will not help Eve in successfully forging a tag for a modified message. While this might seem natural, herein lies the first challenge. In order to use the non-malleability of the NMC, the tampering done by Eve must look like a split-state tampering. If the two states of the non-malleable code are sent directly, the tampering of at least one of them would be dependent on the other, and hence will not be a split-state tampering. Hence, we must find a way to capture this tampering in the interactive setting as a split-state tampering.

To understand how we overcome this challenge, for the sake of simplicity, we will, for now, assume that the adversary is synchronous. Recall that the protocol starts with Alice encoding a MAC key k into (L, R). Since she can't send both simultaneously to Bob (as it would violate split-state tampering), suppose she first sends the state R. The idea then is that Alice will mask R with a one-time pad that she extracts. Specifically, in this modified protocol, Alice initiates the protocol by picking a seed x_R and sending it to Bob. She then uses this seed (as well as her secret w) to extract a mask y_R to hide R. Alice sends this masked string $Z_R = R \oplus Y_R$ to Bob. In the next round, Alice sends the other state L. Finally, Bob uses the received seed in the first step to unmask and get R' and decodes the codeword received to get k'. The main challenge in the security proof is to show that the tampering on L and R can now be captured as two-split-state tamperings. Further, as L is revealed to the adversary, we require the non-malleability to hold, even given the state L. Hence, we require an augmented non-malleable code.

Showing that the above protocol is secure against a synchronous adversary is in itself non-trivial. However, more complications arise when the adversary is asynchronous. Specifically, the order in which the messages are sent to Bob might be altered and hence, the tampering of R itself may end up being dependent on L. To resolve this issue, we borrow the concept of "liveness tests" which was implicit in the protocol due to [RW03] and made explicit in [KR09]. A "Liveness Test" is a two round protocol played between Alice and Bob to ensure that Bob is alive in the protocol. It works as follows: Alice sends the seed to a randomness extractor x as a challenge. Bob is expected to respond with $\mathsf{Ext}(w; x)$. The guarantee, which follows from extractor security, is that if Bob doesn't respond to the liveness test, then Eve can't respond to Alice on her own. It can be used to ensure synchrony in the presence of an asychnronous adversary as follows: at the end of each round from Alice to Bob, Bob will be expected to respond to the liveness test. While this is the high level approach, this interleaving of the liveness test and the choice of the messages sent in each round, needs to be done with care to prevent dependency issues from arising.

Proof technique. The major challenge in the security proof is to capture the tampering made by Eve as a split-state tampering of the two states. In order to justify this, our first step is to prove that Eve is guaranteed to be caught with high probability, if she behaves asynchronously and gains no more advantage than the synchronous setting. We structure the protocol, so that all the useful information is sent by Alice. This means we only have to ensure, through the liveness tests, that Bob remains alive in between any two messages sent by Alice. Once we move into the analysis for the synchronous setting, we wish to use the extractor security to guarantee that Z_R looks uniform and hence the tampering on L can be defined independent of R. While intuitively this looks straight forward, the proof requires a careful analysis of the auxiliary information (which are, for example, the liveness test responses), and a suitable use of extractor security to carefully define the correct tampering functions acting on the two states. In particular, once Z_R is replaced by a uniformly chosen string and not the output of an extractor, how can we ensure that the tampering of R is still consistent with the desired tampering function? We accomplish this by carefully redefining the tampering function acting on R so that it still remains split-state and, at the same time, produces a consistent output as the original tampering function. Once this is done, we use the non-malleability of the underlying NMCs to ensure that the modified key k', if altered, is independent of k. This helps us use the MAC security for the desired robustness.

With the high-level approach described above, we obtain the following theorems:
Informal Theorem: *Assuming the existence of constant rate, optimal error, two-state augmented non-malleable code there exists a* 8*-round privacy amplification protocol with optimal entropy loss and min-entropy requirement* $\Omega(\log(n) + \kappa)$.
Informal Theorem: *Instantiating our construction with the current best known augmented non-malleable code for* 2*-split-state family [Li17], we get a* 8*-round privacy amplification protocol with entropy loss* $\mathcal{O}(\log(n) + \kappa \log(\kappa))$ *and min-entropy requirement* $\Omega(\log(n) + \kappa \log(\kappa))$.

Further, even though the overview above relied on (augmented) Non-malleable Codes, it suffices to use (augmented) Non-malleable Randomness Encoders. As mentioned earlier, there is evidence to suggest that it may be easier to build NMREs with good parameters as opposed to NMCs. Indeed we already know how to build 2-split-state NMREs with constant rate though not augmented. (We do not know of NMCs with constant rate in the 2-split state model.) Additionally, if we were able to make these NMREs augmented as well as with optimal error (i.e., $2^{-\kappa}$), we would be able to obtain a privacy amplification protocol with optimal entropy loss and optimal min-entropy requirement.

Informal Theorem:*Assuming the existence of constant rate, optimal error, two-state augmented non-malleable randomness encoder there exists a 8-round privacy amplification protocol with optimal entropy loss and min-entropy requirement* $\Omega(\log(n) + \kappa)$.

1.4 Organization of the Paper

We explain the preliminaries and the building blocks required for the main protocol in Sections 2 and 3. Then, we explain the construction of the protocol in Section 4. The complete security proofs of the interactive authentication protocol and the privacy amplification protocol are given in Sections 5.1 and 5.2. We also mention an alternate construction of the protocol in Section 5.4 and give a brief proof sketch therein.

2 Preliminaries

Notation. κ denotes security parameter throughout. $s \in_R S$ denotes uniform sampling from set S. $x \leftarrow X$ denotes sampling from a probability distribution X. $x||y$ represents concatenation of two binary strings x and y. $|x|$ denotes length of binary string x. U_l denotes the uniform distribution on $\{0,1\}^l$. All logarithms are base 2.

Statistical distance and Entropy. Let X_1, X_2 be two probability distributions over some set S. Their *statistical distance* is

$$\mathbf{SD}(X_1, X_2) \stackrel{\text{def}}{=} \max_{T \subseteq S}\{\Pr[X_1 \in T] - \Pr[X_2 \in T]\} = \frac{1}{2}\sum_{s \in S}\left|\Pr_{X_1}[s] - \Pr_{X_2}[s]\right|$$

(they are said to be ε-close if $\mathbf{SD}(X_1, X_2) \leq \varepsilon$ and denoted by $X_1 \approx_\varepsilon X_2$). For an event E, $\mathbf{SD}_E(A; B)$ denotes $\mathbf{SD}((A)|E; (B)|E)$
The *min-entropy* of a random variable W is $\mathbf{H}_\infty(W) = -\log(\max_w \Pr[W = w])$. For a joint distribution (W, E), define the (average) conditional min-entropy of W given E [DORS08] as

$$\widetilde{\mathbf{H}}_\infty(W \mid E) = -\log(\mathop{\mathbf{E}}_{e \leftarrow E}(2^{-\mathbf{H}_\infty(W|E=e)}))$$

(here the expectation is taken over e for which $\Pr[E = e]$ is nonzero). For a random variable W over $\{0,1\}^n$, $W|E$ is said to be an (n, t) - source if $\widetilde{\mathbf{H}}_\infty(W|E) \geq t$.
We first prove the following simple lemma:

Lemma 1. *Let A, B be any two independent distributions on $\mathcal{A}, \mathcal{B}$ respectively. Let C be the distribution defined by $C := f(A, B)$ for some deterministic function f. Then, the following distributions will be identical:*

$\mathcal{D}_1$:	$\mathcal{D}_2$:
– $a \leftarrow A$	– $a \leftarrow A$
– $b \leftarrow B$	– $b \leftarrow B$
– $c = f(a, b)$	– $c = f(a, b)$
– *Output* a, b, c	– $a' \leftarrow A \vert f(A, b) = c$
	– *Output* a', b, c

Proof. Let the distributions on $\mathcal{A}, \mathcal{B}, f(\mathcal{A}, \mathcal{B})$ corresponding to the two distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ above be denoted by $(A^{\mathcal{D}_1}, B, C)$ and $(A^{\mathcal{D}_2}, B, C)$ respectively (the distributions on $\mathcal{B}$ and $f(\mathcal{A}, \mathcal{B})$ are identical in $\mathcal{D}_1$ and $\mathcal{D}_2$). Consider

$$\begin{aligned}
\Pr[A^{\mathcal{D}_2} = a, B = b, C = c] &= \Pr[A^{\mathcal{D}_2} = a | B = b, C = c].\Pr[B = b, C = c] \\
&= \Pr[A^{\mathcal{D}_1} = a | f(A^{\mathcal{D}_1}, B) = C, B = b, C = c].\Pr[B = b, C = c] \\
&= \Pr[A^{\mathcal{D}_1} = a | f(A^{\mathcal{D}_1}, B) = c, B = b].\Pr[B = b, C = c] \\
&= \Pr[A^{\mathcal{D}_1} = a | C = c, B = b].\Pr[B = b, C = c] \\
&= \Pr[A^{\mathcal{D}_1} = a, B = b, C = c]
\end{aligned}$$

□

Lemma 2. *For any random variables* A, B, C *if* $(A, B) \approx_\epsilon (A, C)$*, then* $B \approx_\epsilon C$

Lemma 3. *For any random variables* A, B *if* $A \approx_\epsilon B$*, then for any function* f*,* $f(A) \approx_\epsilon f(B)$

Lemma 4. *[DORS08] Let* A, B, C *be random variables. Then*

(a) For any $\delta > 0$*, the conditional entropy* $\mathbf{H}_\infty(A|B = b)$ *is at least* $\widetilde{\mathbf{H}}_\infty(A|B) - \log(1/\delta)$ *with probability at least* $1 - \delta$ *over choice of* b
(b) If B *has at most* 2^λ *possible values, then* $\widetilde{\mathbf{H}}_\infty(A \mid B) \geq \mathbf{H}_\infty(A, B) - \lambda \geq \mathbf{H}_\infty(A) - \lambda$*. and, more generally,* $\widetilde{\mathbf{H}}_\infty(A \mid B, C) \geq \widetilde{\mathbf{H}}_\infty(A, B \mid C) - \lambda \geq \widetilde{\mathbf{H}}_\infty(A \mid C) - \lambda$

Further propositions about statistical distance have been proved in Appendix A.

2.1 Definitions

We define an Interactive Authentication protocol:

Definition 1. *([CKOR10]) An interactive protocol* (A, B) *played by Alice and Bob on a communication channel fully controlled by an adversary* Eve*, is a* (h_W, κ)*-**interactive authentication protocol** if* $\forall m$*, it satisfies the following properties whenever* $\mathbf{H}_\infty(W) \geq h_W$ *and* $m_a = m$*:*

1. *Correctness. If* Eve *is passive,* $\Pr[m_a = m_b] = 1$*.*
2. *Robustness. For any* Eve*, the probability that the following experiment outputs "*Eve *wins" is at most* $2^{-\kappa}$*: sample* w *from* W*; let* $received_a, received_b$ *be the messages received by Alice and Bob upon execution of* (A, B) *with* Eve *actively controlling the channel, and let* $A(w, received_a, r_a, m_a) = t_A$*,* $B(w, received_b, r_b) = (m_b, t_B)$*. Output "*Eve *wins" if* $(m_b \neq m_a \wedge t_B =$ *"accept").*

Further, we define a privacy amplification protocol:

Definition 2. *([CKOR10]) An interactive protocol* (A, B) *played by Alice and Bob on a communication channel fully controlled by an adversary* Eve*, is a* $(h_W, \lambda_k, \delta, \epsilon)$*-**privacy amplification protocol** if it satisfies the following properties whenever* $\mathbf{H}_\infty(W) \geq h_W$*:*

1. *Correctness. If* Eve *is passive,* $\Pr[k_A = k_B] = 1$*.*
2. *Robustness. For any* Eve*, the probability that the following experiment outputs "*Eve *wins" is at most* $2^{-\delta}$*: sample* w *from* W*; let* $received_a, received_b$ *be the messages received by Alice and Bob upon execution of* (A, B) *with* Eve *actively controlling the channel, and let* $A(w, received_a, r_a) = k_A$*,* $B(w, received_b, r_b) = k_B$*. Output "*Eve *wins" if* $(k_A \neq k_B \wedge k_A \neq \perp \wedge k_B \neq \perp)$*.*

3. Extraction. *Define $purify(r)$ to be a randomized function whose input is either a binary string or $\perp$. If $r = \perp$, then $purify(r) = \perp$; else, $purify(r)$ is a uniformly chosen random string of length λ_k. Let $Sent_a, Sent_b$ be the messages sent by Alice and Bob upon execution of (A, B) in presence of* Eve. *Note that the pair $Sent = (Sent_a, Sent_b)$ contains an active* Eve*'s view of the protocol. We require that for any* Eve*,*

$$\mathbf{SD}\left((k_A, Sent), (purify(k_A), Sent)\right) \le \epsilon$$

$$\mathbf{SD}\left((k_B, Sent), (purify(k_B), Sent)\right) \le \epsilon$$

Definition 3. *[KOS18] Let* $(\mathsf{NMREnc}, \mathsf{NMRDec})$ *be s.t.* $\mathsf{NMREnc} : \{0,1\}^r \rightarrow \{0,1\}^k \times (\{0,1\}^{n_1} \times \{0,1\}^{n_2})$ *is defined as* $\mathsf{NMREnc}(r) = (\mathsf{NMREnc}_1(r), \mathsf{NMREnc}_2(r)) = (m, (x, y))$ *and* $\mathsf{NMRDec} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \rightarrow \{0,1\}^k$.
We say that $(\mathsf{NMREnc}, \mathsf{NMRDec})$ *is a* ϵ-***non-malleable randomness encoder*** *with message space* $\{0,1\}^k$ *and codeword space* $\{0,1\}^{n_1} \times \{0,1\}^{n_2}$*, for the distribution* $\mathcal{R}$ *on* $\{0,1\}^r$ *with respect to the 2-split-state family* $\mathcal{F}$ *if the following is satisfied:*

- ***Correctness****:*

$$\Pr_{r \leftarrow \mathcal{R}}[\mathsf{NMRDec}(\mathsf{NMREnc}_2(r)) = \mathsf{NMREnc}_1(r)] = 1$$

- ***Non-malleability****: For each* $(f, g) \in \mathcal{F}$*,* $\exists$ *a distribution* $\mathsf{NMRSim}_{f,g}$ *over* $\{0,1\}^k \cup \{same^*, \perp\}$ *such that*

$$\mathsf{NMRTamper}_{f,g} \approx_\epsilon Copy(U_k, \mathsf{NMRSim}_{f,g})$$

where $\mathsf{NMRTamper}_{f,g}$ *denotes the distribution* $(\mathsf{NMREnc}_1(\mathcal{R}), \mathsf{NMRDec}((f,g)(\mathsf{NMREnc}_2(\mathcal{R})))$[2] *and* $Copy(U_k, \mathsf{NMRSim}_{f,g})$ *is defined as:*

$$u \leftarrow U_k;\ \tilde{m} \leftarrow \mathsf{NMRSim}_{f,g}$$

$$Copy(u, \tilde{m}) = \begin{cases} (u, u), & \text{if } \tilde{m} = same^* \\ (u, \tilde{m}), & \text{otherwise} \end{cases}$$

$\mathsf{NMRSim}_{f,g}$ *should be efficiently samplable given oracle access to* $(f, g)(.)$.

Further, the rate of this code is defined as $k/(n_1 + n_2)$

Definition 4. *Let* $(\mathsf{NMREnc}, \mathsf{NMRDec})$ *be s.t.* $\mathsf{NMREnc} : \{0,1\}^r \rightarrow \{0,1\}^k \times (\{0,1\}^{n_1} \times \{0,1\}^{n_2})$ *is defined as* $\mathsf{NMREnc}(r) = (\mathsf{NMREnc}_1(r), \mathsf{NMREnc}_2(r)) = (m, (x, y))$ *and* $\mathsf{NMRDec} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \rightarrow \{0,1\}^k$.
We say that $(\mathsf{NMREnc}, \mathsf{NMRDec})$ *is a* ϵ-***augmented non-malleable randomness encoder*** *with message space* $\{0,1\}^k$ *and codeword space* $\{0,1\}^{n_1} \times \{0,1\}^{n_2}$*, for the distribution* $\mathcal{R}$ *on* $\{0,1\}^r$ *with respect to the 2-split-state family* $\mathcal{F}$ *if the following is satisfied:*

[2] Here $(f, g)(\mathsf{NMREnc}_2(\mathcal{R}))$ just denotes the tampering by the split-state tampering functions f and g on the corresponding states.

– ***Correctness:***

$$\Pr_{r \leftarrow \mathcal{R}}[\mathsf{NMRDec}(\mathsf{NMREnc}_2(r)) = \mathsf{NMREnc}_1(r)] = 1$$

– ***Non-malleability:*** *For each* $(f, g) \in \mathcal{F}$, $\exists$ *a distribution* $\mathsf{NMRSim}_{f,g}$ *over* $\{0,1\}^{n_1} \times \{\{0,1\}^k \cup \{same^*, \perp\}\}$ *such that*

$$\mathsf{NMRTamper}^+_{f,g} \approx_\epsilon Copy(U_k, \mathsf{NMRSim}^+_{f,g})$$

where $\mathsf{NMRTamper}^+_{f,g}$ *denotes the distribution* $(\mathsf{NMREnc}_1(\mathcal{R}), L, \mathsf{NMRDec}((f(L), g(R)))$ *where* $(L, R) \equiv \mathsf{NMREnc}_2(\mathcal{R})$ *and* $Copy(U_k, \mathsf{NMRSim}^+_{f,g})$ *is defined as:*

$$u \leftarrow U_k;\ L, \tilde{m} \leftarrow \mathsf{NMRSim}^+_{f,g}$$

$$Copy(u, \tilde{m}) = \begin{cases} (u, L, u), & \text{if } \tilde{m} = same^* \\ (u, L, \tilde{m}), & \text{otherwise} \end{cases}$$

$\mathsf{NMRSim}^+_{f,g}$ *should be efficiently samplable given oracle access to* $(f, g)(.)$.

3 Buliding Blocks

We use information-theoretic message authentication codes, strong average case extractor and an augmented non-malleable code for 2-split-state family , as building blocks to our construction. We define these building blocks below.

3.1 Augmented Non-malleable Codes

We first define Augmented Non-malleable codes for the 2-split-state family as below:

Definition 5. ***Augmented Non-malleable Codes*** *[AAG+16] A coding scheme* $(\mathsf{Enc}, \mathsf{Dec})$ *with message and codeword spaces as* $\{0,1\}^\alpha, (\{0,1\}^\beta)^2$ *respectively, is* ϵ*- augmented-non-malleable with respect to the function family* $\mathcal{F} = \{(f_1, f_2) : f_i : \{0,1\}^\beta \to \{0,1\}^\beta\}$ *if* $\forall$ $(f_1, f_2) \in \mathcal{F}$, $\exists$ *a distribution* Sim_{f_1,f_2} *over* $(\{0,1\}^\beta) \times (\{0,1\}^\alpha \cup \{same^*, \perp\})$ *such that* $\forall\, m \in \{0,1\}^\alpha$

$$\mathsf{Tamper}^m_{f_1,f_2} \approx_\epsilon \mathsf{Copy}^m_{Sim_{f_1,f_2}}$$

where $\mathsf{Tamper}^m_{f_1,f_2}$ *denotes the distribution* $(L, \mathsf{Dec}(f_1(L), f_2(R)))$, *where* $\mathsf{Enc}(m) = (L, R)$. $\mathsf{Copy}^m_{Sim_{f_1,f_2}}$ *is defined as*

$$(L, \tilde{m}) \leftarrow Sim_{f_1,f_2}$$

$$\mathsf{Copy}^m_{Sim_{f_1,f_2}} = \begin{cases} (L, m)\ if\ (L, \tilde{m}) = (L, same^*) \\ (L, \tilde{m}) \text{ otherwise} \end{cases}$$

Sim_{f_1,f_2} should be efficiently samplable given oracle access to $(f_1, f_2)(.)$. [3] *We say an ϵ- augmented non-malleable code has optimal error, if $\epsilon \le 2^{-\Theta(\alpha)}$. We express the rate, of an augmented non-malleable code as a function of α. We say the rate is a function $r(.)$, if $2\beta = (\alpha/r(\alpha))$ i.e codeword length $= \frac{message\ length}{r(message\ length)}$. Similarly, the ϵ-non-malleable code has error $2^{-\phi(.)}$, if $\epsilon \le 2^{-\phi(.)}$*

3.2 Information-theoretic One-Time Message Authentication Codes

A family of pair of functions $\{\mathsf{Tag}_{k_a} : \{0,1\}^\gamma \to \{0,1\}^\delta,\ \mathsf{Vrfy}_{k_a} : \{0,1\}^\gamma \times \{0,1\}^\delta \to \{0,1\}\}_{k_a \in \{0,1\}^\tau}$ is said to a μ − secure one time MAC if

1. For $k_a \in_R \{0,1\}^\tau$, $\forall\, m \in \{0,1\}^\gamma$, $\Pr[\mathsf{Vrfy}_{k_a}(m, \mathsf{Tag}_{k_a}(m)) = 1] = 1$
2. For any $m \neq m', t, t'$, $\Pr_{k_a}[\mathsf{Tag}_{k_a}(m) = t | \mathsf{Tag}_{k_a}(m') = t'] \le \mu$ for $k_a \in_R \{0,1\}^\tau$

3.3 Average-case Extractors

Definition 6. *[DORS08, Section 2.5] Let $Ext : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^l$ be a polynomial time computable function. We say that Ext is an efficient average-case (n, t, d, l, ϵ)-strong extractor if for all pairs of random variables (W, I) such that W is an n-bit string satisfying $\widetilde{\mathbf{H}}_\infty(W|I) \ge t$, we have $\mathbf{SD}\left((Ext(W; X), X, I), (U, X, I)\right) \le \epsilon$, where X is uniform on $\{0,1\}^d$.*

We now prove the following lemma about strong extractors.

Lemma 5. *Let W be a source with min-entropy t and Ext be an (n, t, d, l, ϵ)-strong extractor.Then the following distributions are 1.5ϵ-close.*

− $x \in_R \{0,1\}^d$ − $w \leftarrow W$ − $y = \mathsf{Ext}(w; x)$ − *Output* x, y, w	− $x \in_R \{0,1\}^d$ − $y' \in_R \{0,1\}^l$ − *If* $\exists\ w^{res} \in Support(W)$, *such that* $y, = \mathsf{Ext}(w; x)$ $w^{res} \leftarrow W|\mathsf{Ext}(W; x) = y'$ *else* $w^{res} = \perp$ − *Output* x, y', w^{res}

Proof. $2\mathbf{SD}\left((X, Y, W); (X, Y', W^{res})\right)$

$$= \underbrace{\sum_{x,y} \Pr[X = x, Y' = y, W^{res} = \perp]}_{*}$$
$$+ \underbrace{\sum_{x,y,w} |\Pr[X = x, Y = y, W = w] - \Pr[X = x, Y' = y, W^{res} = w]|}_{**}$$

[3] For simplicity in the proof, we may assume here that the decoder Dec never outputs $\perp$. This can be done by replacing $\perp$ with some fixed string, like 00..0.

Let $S(W,X)$ denote the set $\{\mathsf{Ext}(w;x): \ w \in Support(W) \wedge x \in \{0,1\}^d\}$.

$$\begin{aligned}\sum_{x,y} \Pr[X=x, Y'=y, W^{res}=\perp] &= \Pr[W^{res}=\perp] \\ &= \Pr[Y' \notin S(W,X)] \\ &\leq \Pr[\mathsf{Ext}(W;X) \notin S(W,X)] + \mathbf{SD}\left(Y'; \mathsf{Ext}(W;X)\right) \\ &\leq 0 + \epsilon = \epsilon\end{aligned}$$

The above inequality follows because W is a t-source and Ext is a strong extractor. Hence Observe that for any $(x,y,w(\neq \perp))$, $\Pr[X=x, Y=y, W=w]=0$ iff $\Pr[X=x, Y'=y, W^{res}=w]=0$. So to compute $**$ we only need to consider (x,y,w) such that $\Pr[X=x, Y=y, W=w]>0$. For any (x,y,w) such that $\Pr[X=x, Y=y, W=w]>0$,

$$\begin{aligned}\Pr[X=x, Y'=y, W^{res}=w] &= \Pr[W^{res}=w | X=x, Y'=y] \Pr[X=x, Y'=y] \\ &= \Pr[W=w | \mathsf{Ext}(W;X)=Y', X=x, Y'=y] \Pr[X=x, Y'=y] \\ &= \Pr[W=w | \mathsf{Ext}(W;x)=y] \Pr[X=x, Y'=y]\end{aligned}$$

The above equation follows because the event $W=w$ is independent of the event $(X=x, Y'=y)$ and $\Pr[\mathsf{Ext}(W;x)=y]>0$ (as $\Pr[X=x, Y=y, W=w]>0$).

$$\begin{aligned}&\sum_{x,y,w} |\Pr[X=x, Y=y, W=w] - \Pr[X=x, Y'=y, W^{res}=w]| \\ &= \sum_{x,y,w} \Pr[W=w|\mathsf{Ext}(W;x)=y].|\Pr[X=x, Y=y] - \Pr[X=x, Y'=y]| \\ &= \sum_{x,y} |\Pr[X=x, Y=y] - \Pr[X=x, Y'=y]| \sum_{w} \Pr[W=w|\mathsf{Ext}(W;x)=y] \\ &= \sum_{x,y} |\Pr[X=x, Y=y] - \Pr[X=x, Y'=y]|.1 \qquad (1) \\ &\leq 2\epsilon\end{aligned}$$

Equation 1 follows from Proposition 3 (Appendix A). Let A_w denote the event $(W=w)$. Let B be the event $(\mathsf{Ext}(W;x)=y)$ and the mutually exclusive and exhaustive events are $\{A_w\}_{w\in\{0,1\}^n}$. B is a non-zero probability event as $\Pr[\mathsf{Ext}(W;x)=y]>0$. Equation 1 follows by applying Proposition 3(Appendix A) to $\{A_w\}_{w\in\{0,1\}^n}$ and B.
Therefore we have

$$\mathbf{SD}\left((X,Y,W);(X,Y',W^{res})\right) \leq 1.5\epsilon$$

We will state two useful inequalities from the proof of Lemma 5 that we will further use in the paper.

$$\Pr[W^{res}=\perp] \leq \epsilon \qquad (2)$$

$$\Pr[W^{res}\neq\perp].\mathbf{SD}_{W^{res}\neq\perp}(X,Y,W);(X,Y',W^{res}) \leq 0.5\epsilon \leq \epsilon \qquad (3)$$

Inequality 3 follows from Proposition 1(Appendix A) and Lemma 5. □

We also require the following lemma given in [CKOR10]:

Lemma 6. *Let* Ext *be a* (n, t, d, l, ϵ)*-strong extractor and* W *be a random variable over* $\{0,1\}^n$ *with* $\mathbf{H}_\infty(W) \geq t$*. Then* $\Pr_x[\mathbf{H}_\infty(\mathsf{Ext}(W; x)) \leq l - 1] \leq 2^l \epsilon$.

4 Protocol

4.1 Notation

- Let Ext' be an $(n, t', d, 3l', \epsilon_1)$- average case extractor.
- Let Ext be an (n, t, d, l, ϵ_2)- average case extractor.
- Let $\mathsf{Enc}, \mathsf{Dec}$ be an ϵ_3- secure two-state augmented non-malleable code with message, codeword spaces being $\{0,1\}^\tau$ and $\{0,1\}^{2l}$.
- Let $\mathsf{Tag}, \mathsf{Vrfy}$ be an ϵ_4-secure one-time MAC with key, message and tag spaces being $\{0,1\}^\tau$, $\{0,1\}^d$, and $\{0,1\}^\delta$ respectively.
- Let Ext'' be an $(n, t'', d, l'', \epsilon_5)$- average case extractor.

4.2 Protocol

We now describe the Privacy Amplification Protocol below. w is drawn from the entropic source W, and is shared between Alice and Bob. We denote the Interactive Authentication Protocol to authenticate a message m by $\pi^{\mathsf{AUTH}}_{m,w}$ and the Privacy Amplification Protocol by π^{PA}_w.

As described in the introduction, the idea behind the protocol is as follows: For the synchronous setting: Alice picks a MAC key, encodes it using the NMC and sends across the states to Bob. Now, in order to ensure that the tampering done by Eve is captured as a split-state tampering on the states, Alice uses an extractor and masks one of the states before sending it. In the next round, the other state is sent in clear. We require the augmented nature of the NMC to guarantee security even when one state is sent in clear. For the asynchronous setting, we need to add "liveness tests" to the protocol (where an extractor seed is sent by one party as a challenge and the other party has to respond to this correctly). By the nature of the protocol, as the communication is unidirectional(all the "useful information" is only sent by Alice), we only need to include liveness tests to ensure that Bob is alive. For this, Alice sends a liveness test seed for a long extractor output in the first step. This challenge seed is reused for the liveness test responses. The reuse of liveness test seed reduces the number of rounds in the protocol. But, in addition, it is also crucial that this is done to guarantee security of protocol, else dependencies arise.

Theorem 1 *Let* $(\mathsf{Enc}, \mathsf{Dec})$, $(\mathsf{Tag}, \mathsf{Vrfy})$, Ext' *and* Ext *be as in Section 4.1. Then, the 8-round sub-protocol* π^{AUTH} *in Figure 1 is a an* (t', κ)*-interactive message authentication protocol.*

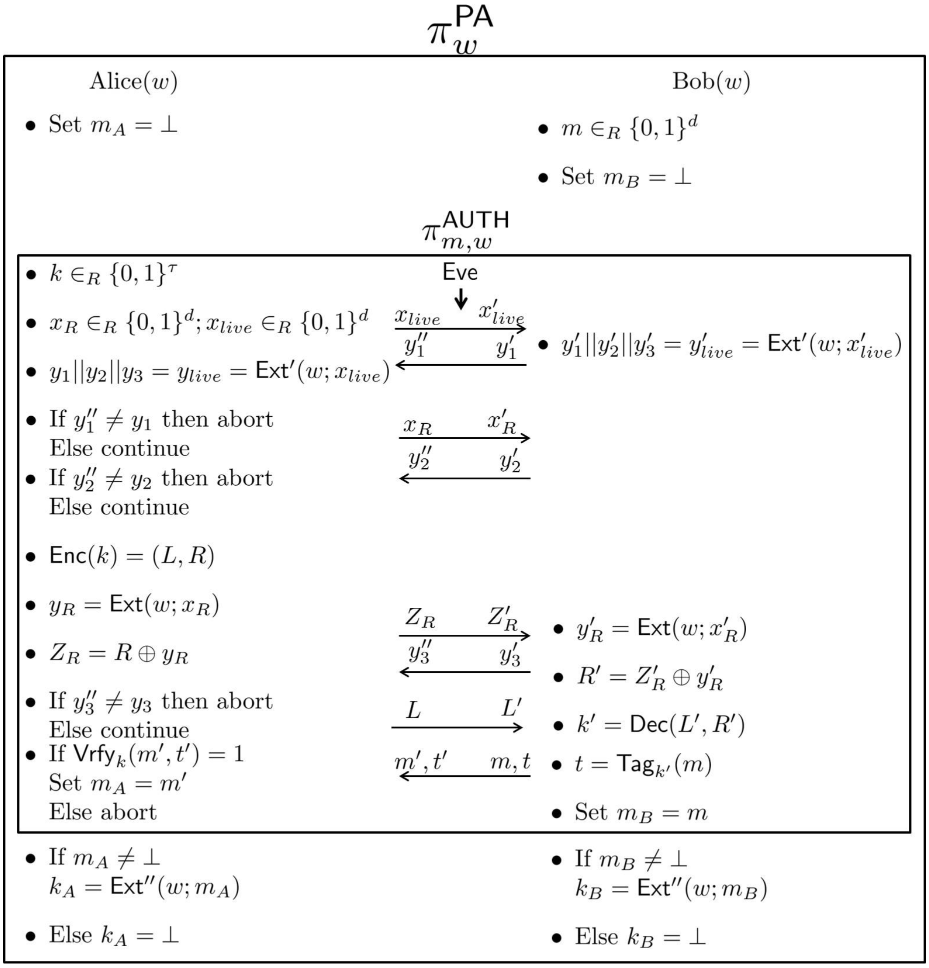


Fig. 1. Privacy Amplification Protocol

Theorem 2.A *If* $(\mathsf{Enc}, \mathsf{Dec})$ *in* π^{AUTH} *is a two-state, constant rate, optimal error augmented non-malleable code, then the 8-round protocol* π^{PA} *in Figure 1 is a* $(t', l'', \kappa, \kappa + 1)$*-secure privacy amplification protocol with optimal entropy loss and with min-entropy requirement* $\Omega(\log(n) + \kappa)$.

Theorem 2.B *If* $(\mathsf{Enc}, \mathsf{Dec})$ *in* π^{AUTH} *is instantiated with the augmented non-malleable code given in [Li17], then the 8-round protocol* π^{PA} *in Figure 1 is a* $(t', l'', \kappa, \kappa + 1)$*-secure privacy amplification protocol with entropy loss being* $\mathcal{O}(\log(n) + \kappa log(\kappa))$ *and with min-entropy requirement* $\Omega(\log(n) + \kappa \log \kappa)$.

5 Security proof

5.1 Proof of Theorem 1

We first prove that π^{AUTH} is an interactive authentication protocol.
Correctness: The correctness of π^{AUTH} follows easily.
Robustness: We need to show that

$$\Pr[\mathsf{Eve}\text{ wins}] = \Pr[m_A \neq m_B \wedge m_A \neq \perp \wedge m_B \neq \perp] \leq 2^{-\kappa}.$$

If Bob didn't receive any messages during the protocol, then $m_B = \perp$ and Eve doesn't win. Further, for Eve to win, all the liveness test checks must have verified correctly. Hence, from now on, we assume Bob receives and sends messages and that the liveness test checks go through. We now analyze Eve's success probability by considering the asynchronous and synchronous case separately. We define the following events for the same.

- Let Sync denote the event that Eve is synchronous and doesn't interleave.
- Async denote the complement of the event Sync, i.e., where Eve interleaves.
- Pass denote the event that Eve passes all initial checks done by Alice and Bob. It denotes the event "$(y_1''||y_2''||y_3'' = y_{live})$". Pass also implies $m_B = m \neq \perp$.

Then, we get:

$$\begin{aligned}\Pr[\mathsf{Eve}\text{ wins}] &\leq \Pr[\mathsf{Eve}\text{ wins}|\mathsf{Sync}] + \Pr[\mathsf{Eve}\text{ wins}|\mathsf{Async}] \\ &\leq \Pr[\mathsf{Eve}\text{ wins}|\mathsf{Sync}, \mathsf{Pass}] + \Pr[\mathsf{Eve}\text{ wins}|\mathsf{Async}] \qquad (4)\end{aligned}$$

This is because, the event Eve wins implies that Pass has occured. To prove robustness we will now bound each of the above summands.

Lemma 7. $\Pr[\mathsf{Eve}\ wins|\mathsf{Async}] \leq \Pr[\mathsf{Eve}\ wins|\mathsf{Sync}, \mathsf{Pass}] + 2.(2^{-l'} + 2^{3l'}\epsilon_1 + 2^{-\lambda})$

Proof. We first introduce the following notations:

- Let msg_i denote the message received by Eve in the actual i-th round (i.e., in the synchronous world) of the protocol ($msg_1 = x_{live}$, $msg_2 = y_1', \cdots, msg_8 = m, t$).
- Let msg_i' denote the modification of msg_i sent by Eve to Bob/Alice ($msg_1' = x_{live}'$, $msg_2' = y_1'', \cdots, msg_8' = m', t'$. In the asynchronous setting these modified messages may depend on messages received by Eve in later rounds).

We split the event Async into the following three mutually exclusive events:

- Indlive: This denotes the event that Eve sends x_{live}' to Bob before she receives x_{live} from Alice.
- $\mathsf{Async}_\mathsf{A}$: This denotes the event that $\mathsf{Indlive}^C$ (the compliment event) occurred and that for some i, Eve receives both msg_i and msg_{i+2} from Alice before she sends msg_i' to Bob.

– $\mathsf{Async_B}$: This denotes the event that $\mathsf{Indlive}^C$ and $\mathsf{Aysnc_A}^C$ occurred and for some i, Eve receives both msg_i and msg_{i+2} from Bob before she sends msg_i' to Alice.

These events are clearly mutually exclusive. Further, observe that $\mathsf{Async} = \mathsf{Indlive} \bigcup \mathsf{Async_A} \bigcup \mathsf{Async_B}$. This is because, given that $\mathsf{Async_A}^C$ and $\mathsf{Indlive}^C$ occurred, as we are in the asynchronous setting, the additional (third) condition described in $\mathsf{Async_B}$ must occur. Then, we have:

$$\Pr[\mathsf{Eve}\text{ wins}|\mathsf{Async}] \leq \Pr[\mathsf{Eve}\text{ wins}|\mathsf{Indlive}] + \Pr[\mathsf{Eve}\text{ wins}|\mathsf{Async_A}] + \Pr[\mathsf{Eve}\text{ wins}|\mathsf{Async_B}] \tag{5}$$

We now bound each of the summands above separately

Claim 1 $\Pr[\mathsf{Eve}\ wins|\mathsf{Indlive}] \leq 2^{-l'} + 2^{3l'}.\epsilon_1 + 2^{-\lambda}$

Proof. In this case, as the liveness test seed sent by Eve is independent of X_{live} (as x'_{live} is sent before receiving x_{live} from Alice), the liveness test responses seen by Eve are independent of X_{live}. Hence, to win the game, Eve has to respond to at least one liveness test challenge correctly on her own. Conditioned on the event that the liveness test seed is good so that the extractor output on x_{live} has high entropy, we can show that Eve can do no better than guess the liveness test response. Let the part of transcript seen by Eve until she responds to this liveness test on her own be denoted by E. Then, E is an auxiliary information, independent of seed X_{live}. By setting parameters we can ensure that $\widetilde{\mathbf{H}}_\infty(W|E) \geq t' + \lambda$, so that by Lemma 4, E takes a value such that $\mathbf{H}_\infty(W|E=e) \geq t'$ with probability at least $1 - 2^{-\lambda}$. For value of E such that $\mathbf{H}_\infty(W|E=e) \geq t'$, by applying Lemma 6 we get:

$$\begin{aligned}
\Pr[\mathsf{Eve}\text{ wins}|\mathsf{Indlive}] &\leq \Pr[\mathsf{Pass}|\mathsf{Indlive}] \\
&\leq \Pr[\mathsf{Pass}|\mathsf{Indlive}, \mathbf{H}_\infty(\mathsf{Ext}'(W; x_{live})|E=e) > l' - 1] \\
&+ \Pr_{x_{live}}[\mathbf{H}_\infty(\mathsf{Ext}'(W; x_{live})|E=e) \leq l' - 1|\mathsf{Indlive}] \\
&\leq \Pr[\mathsf{Pass}|\mathsf{Indlive}, \mathbf{H}_\infty(\mathsf{Ext}'(W; x_{live})|E=e) > l' - 1] \\
&+ \Pr_{x_{live}}[\mathbf{H}_\infty(\mathsf{Ext}'(W; x_{live})|E=e) \leq l' - 1|\mathsf{Indlive}, \mathbf{H}_\infty(W|E=e) \geq t'] \\
&+ \Pr_e[\mathbf{H}_\infty(W|E=e) < t'|\mathsf{Indlive}] \\
&\leq 2^{-l'} + 2^{3l'}\epsilon_1 + 2^{-\lambda}
\end{aligned} \tag{6}$$

Claim 2 $\Pr[\mathsf{Eve}\ wins|\mathsf{Async_A}] \leq 2^{-l'} + \epsilon_1 2^{3l'} + 2^{-\lambda}$

Proof. Observe that if $\mathsf{Async_A}$ has occurred, it means that Eve would have responded to atleast one of the liveness test responses correctly on her own. Then, conditioned on the event that the liveness test seed is good, i.e., the extractor output on x_{live} has high entropy, we know that Eve can do no better than guessing the liveness test response. As Eve may see certain part of the transcript before she responds to liveness test, as in previous claim, let E denote the part of transcript seen by Eve until she responds a certain liveness test on her own. By similar arguments as in Claim 1, for choice of E such that $\mathbf{H}_\infty(W|E=e) \geq t'$, we apply Lemma 6 (which guarantees that the extractor output has high enough entropy with high probability) to get:

$$\begin{aligned}
\Pr[\mathsf{Eve}\ \text{wins}|\mathsf{Async_A}] &\leq \Pr[\mathsf{Pass}|\mathsf{Async_A}] \\
&\leq \Pr[\mathsf{Pass}|\mathsf{Async_A}, \mathbf{H}_\infty(\mathsf{Ext}'(W;x_{live})|E=e) > l'-1] \\
&+ \Pr_{x_{live}}[\mathbf{H}_\infty(\mathsf{Ext}'(W;x_{live})|E=e) \leq l'-1|\mathsf{Async_A}] \\
&\leq 2^{-l'} + 2^{3l'}.\epsilon_1 + 2^{-\lambda} \qquad (7)
\end{aligned}$$

where the last inequality uses Lemma 6 and Lemma 4, just as in the previous claim.

Claim 3 $\Pr[\mathsf{Eve}\ \mathit{wins}|\mathsf{Async_B}] \leq \Pr[\mathsf{Eve}\ \mathit{wins}|\mathsf{Sync}, \mathsf{Pass}]$

Proof. We aim to prove that given $\mathsf{Aysnc_B}$, Eve only gains as much advantage in winning, as in the synchronous setting. To prove this, we first define the function family $\mathcal{F}_{asyncB}$, which captures the modifications made (to the transcript) by Eve given $\mathsf{Async_B}$. Then, we will prove that for any tampering made by Eve using $(f_1, \cdots, f_8) \in \mathcal{F}_{asyncB}$, we can capture it by a function in the synchronous setting (post removing the liveness test checks, which can only give more advantage to Eve).[4].This would prove that the probability of Eve winning given $\mathsf{Async_B}$ is at most the probability of her winning given $(\mathsf{Sync}, \mathsf{Pass})$. Here, by slight abuse of notation, we also use Pass to denote that the liveness test checks are removed. Now, we define the domain and range spaces of functions $(f_1, \cdots, f_8)$ from $\mathcal{F}_{asyncB}$:

- $f_1 : \{0,1\}^d \to \{0,1\}^d$
- $f_2 : \{0,1\}^d \times \{0,1\}^{l'} \times (\{0,1\}^{l'} \bigcup \{\perp\}) \times (\{0,1\}^{l'} \bigcup \{\perp\}) \times (\{0,1\}^{d+\delta} \bigcup \{\perp\}) \to \{0,1\}^{l'}$
- $f_3 : \{0,1\}^d \times \{0,1\}^{l'} \times (\{0,1\}^d \bigcup \{\perp\}) \to \{0,1\}^d$
- $f_4 : \{0,1\}^d \times \{0,1\}^{l'} \times \{0,1\}^d \times \{0,1\}^{l'} \times (\{0,1\}^{l'} \bigcup \{\perp\}) \times (\{0,1\}^{d+\delta} \bigcup \{\perp\}) \to \{0,1\}^{l'}$
- $f_5 : \{0,1\}^d \times \{0,1\}^{l'} \times \{0,1\}^d \times \{0,1\}^{l'} \times (\{0,1\}^l \bigcup \{\perp\}) \to \{0,1\}^l$
- $f_6 : \{0,1\}^d \times \{0,1\}^{l'} \times \{0,1\}^d \times \{0,1\}^{l'} \times \{0,1\}^l \times \{0,1\}^{l'} \times (\{0,1\}^{d+\delta} \bigcup \{\perp\}) \to \{0,1\}^{l'}$
- $f_7 : \{0,1\}^d \times \{0,1\}^{l'} \times \{0,1\}^d \times \{0,1\}^{l'} \times \{0,1\}^l \times \{0,1\}^{l'} \times (\{0,1\}^l \bigcup \{\perp\}) \to \{0,1\}^l$
- $f_8 : \{0,1\}^d \times \{0,1\}^{l'} \times \{0,1\}^d \times \{0,1\}^{l'} \times \{0,1\}^l \times \{0,1\}^{l'} \times \{0,1\}^l \times \{0,1\}^{d+\delta} \to \{0,1\}^{d+\delta}$

[4] As Eve is information theoretic, we can assume that Eve gives the functions she is going to use for modifications a priori

The $\perp$ symbol is used to denote that the function description does not depend on that input. Given $\mathsf{Async_B}$, we know that $\mathsf{Indlive}^C$ and $\mathsf{Async_A}^C$ have occurred. This means that x'_{live} is sent by Eve only after she sees x_{live}. Further, each of x'_{live}, x'_R, Z'_R are sent by Eve to Bob before she receives the subsequent messages x_R, Z_R, L, respectively, from Alice. Hence the function description of f_1 is exactly as in the synchronous setting and that of f_3, f_5, f_7 only differs from their synchronous counterpart in that, these functions may not depend on certain messages in their input (we use $\perp$ to denote this). The function description f_8 is exactly as in the synchronous setting.
Now, if we assume that Pass occurs and hence remove the liveness test checks in the game, then clearly, it can only increase the advantage of Eve in winning. Hence

$$\Pr[\mathsf{Eve}\text{ wins}|\mathsf{Async_B}] \leq \Pr[\mathsf{Eve}\text{ wins}|\mathsf{Async_B}, \mathsf{Pass}]$$

Two key observations below will complete the proof of this claim:

1. Post the liveness test checks are removed (both in case of asynchronous and synchronous), the functions f_2, f_4 and f_6, which give modifications of the liveness test responses by Bob, are no longer used in generating the view of Eve. So, given that Pass occurred, the view of Eve in both the asynchronous and synchronous world does not depend on the functions f_2, f_4 and f_6.
2. The current descriptions of f_3, f_5, f_7 in $\mathcal{F}_{asyncB}$ can be captured by defining functions f'_3, f'_5, f'_7 (whose domains do not include $\perp$) such that if a certain input for f_i is $\perp$, replace it with a dummy string and use f'_i to get the modification. If all inputs of f_i are $\neq \perp$, f'_i is same as f_i.The function descriptions of f'_3, f'_5, f'_7 are as in the synchronous setting.

Observations 1. and 2. above show that post removing the liveness test checks, i.e., assuming Pass occurred, the function descriptions of $f_1, f_3, f_5, f_7.f_8$ can be captured by function descriptions in the synchronous setting. Hence, it follows that:

$$\begin{aligned}\Pr[\mathsf{Eve}\text{ wins}|\mathsf{Async_B}] &\leq \Pr[\mathsf{Eve}\text{ wins}|\mathsf{Async_B}, \mathsf{Pass}] \\ &\leq \Pr[\mathsf{Eve}\text{ wins}|\mathsf{Sync}, \mathsf{Pass}] \end{aligned} \tag{8}$$

Combining the above claims 1, 2 and 3 in Equation 5, we get:

$$\Pr[\mathsf{Eve}\text{ wins}|\mathsf{Async}] \leq \Pr[\mathsf{Eve}\text{ wins}|\mathsf{Sync}, \mathsf{Pass}] + 2.(2^{-l'} + 2^{3l'}\epsilon_1 + 2^{-\lambda})$$

□

Lemma 8. $\Pr[\mathsf{Eve}\ wins|\mathsf{Sync}, \mathsf{Pass}] \leq 2^{-\lambda} + 3\epsilon_2 + \epsilon_3 + \epsilon_4$

Proof. Let us define the random variable corresponding to the view of Eve conditioned on the event Sync happening. We introduce the following notations for that.

- Let m denote the message being authenticated by Bob through π^{AUTH}.
- As Eve is information theoretic adversary, we assume that she chooses the tampering functions of each round apriori. We denote these functions with the literals $f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8$ with domains and ranges set as follows.
  - $f_1 : \{0,1\}^d \to \{0,1\}^d$
  - $f_2 : \{0,1\}^d \times \{0,1\}^{l'} \to \{0,1\}^{l'}$
  - $f_3 : \{0,1\}^d \times \{0,1\}^{l'} \times \{0,1\}^d \to \{0,1\}^d$
  - $f_4 : \{0,1\}^d \times \{0,1\}^{l'} \times \{0,1\}^d \times \{0,1\}^{l'} \to \{0,1\}^{l'}$
  - $f_5 : \{0,1\}^d \times \{0,1\}^{l'} \times \{0,1\}^d \times \{0,1\}^{l'} \times \{0,1\}^l \to \{0,1\}^l$
  - $f_6 : \{0,1\}^d \times \{0,1\}^{l'} \times \{0,1\}^d \times \{0,1\}^{l'} \times \{0,1\}^l \times \{0,1\}^{l'} \to \{0,1\}^{l'}$
  - $f_7 : \{0,1\}^d \times \{0,1\}^{l'} \times \{0,1\}^d \times \{0,1\}^{l'} \times \{0,1\}^l \times \{0,1\}^{l'} \times \{0,1\}^l \to \{0,1\}^l$
  - $f_8 : \{0,1\}^d \times \{0,1\}^{l'} \times \{0,1\}^d \times \{0,1\}^{l'} \times \{0,1\}^l \times \{0,1\}^{l'} \times \{0,1\}^l \times \{0,1\}^{d+\delta} \to \{0,1\}^{d+\delta}$

$\underline{View0^m_{f_1,..,f_8}}$:

- $w \leftarrow W$
- $x_{live} \in_R \{0,1\}^d$, $k \in_R \{0,1\}^d, x_R \in_R \{0,1\}^d$
- $x'_{live} = f_1(x_{live})$
- $y_1||y_2||y_3 = y_{live} = \mathsf{Ext}'(w; x_{live})$
- $y'_1||y'_2||y'_3 = y'_{live} = \mathsf{Ext}'(w; x'_{live})$
- $x'_R = f_3(x_{live}, y'_1, x_R)$
- $y_R = \mathsf{Ext}(w; x_R)$
- $(L, R) \leftarrow \mathsf{Enc}(k)$
- $z_R = y_R \oplus R$
- $z'_R = f_5(x_{live}, y'_1, x_R, y'_2, z_R)$
- $y'_R = \mathsf{Ext}(w; x'_R)$
- $R' = z'_R \oplus y'_R$
- $L' = f_7(x_{live}, y'_1, x_R, y'_2, z_R, y'_3, L)$
- $k' = \mathsf{Dec}(L', R')$
- $t = \mathsf{Tag}_{k'}(m)$
- $(m', t') = f_8(x_{live}, y'_1, x_R, y'_2, z_R, y'_3, L, m, t)$
- Output $x_{live}, y'_{live}, x_R, z_R, L, m, t$

The random variable $View0^m_{f_1,..,f_8}$ is defined as

$$View0^m_{f_1,..,f_8} \equiv (X_{live}, Y'_{live}, X_R, Z_R, L, m, T)$$

where the capital letters on the right denote the distributions corresponding to the respective small letters (as described in the figure above). Then, we have:

$$\begin{aligned}&\Pr[\mathsf{Eve\ wins}|\mathsf{Sync},\mathsf{Pass}]\\&=\Pr[(m',t')\leftarrow\mathsf{Eve}(View0^m_{f_1,\cdots,f_8})\wedge m'\neq m\wedge\mathsf{Vrfy}_K(m',t')=1]\end{aligned}\qquad(9)$$

where the probability is over the randomness used to generate $View0^m_{f_1,\cdots,f_8}$, namely W, X_{live}, K, X_R and the randomness used in Enc. To bound this probability, we use a hybrid argument. We now define the views of Eve in the subsequent hybrids. Then, we prove that the success probability of Eve given $View0^m_{f_1,..,f_8}$ (Equation 9) is upper bounded by its success probability given the final view ($View4^m_{f_1,..,f_8}$) upto a small error.

$View1^m_{f_1,..,f_8}$:

- $w \leftarrow W$
- $x_{live} \in_R \{0,1\}^d$, $k \in_R \{0,1\}^d, x_R \in_R \{0,1\}^d$
- $x'_{live} = f_1(x_{live})$
- $y_1||y_2||y_3 = y_{live} = \mathsf{Ext}'(w; x_{live})$
- $y'_1||y'_2||y'_3 = y'_{live} = \mathsf{Ext}'(w; x'_{live})$
- $\tilde{w} \leftarrow \tilde{W}$
  $\tilde{W}$ is the conditional source, $\tilde{W} := W|(\mathsf{Ext}'(W; x_{live}) = y_{live}, \mathsf{Ext}'(W; x'_{live}) = y'_{live})$
- $x'_R = f_3(x_{live}, y'_1, x_R)$
- $y_R = \mathsf{Ext}(\tilde{w}; x_R)$
- $(L, R) \leftarrow \mathsf{Enc}(k)$
- $z_R = y_R \oplus R$
- $z'_R = f_5(x_{live}, y'_1, x_R, y'_2, z_R)$
- $y'_R = \mathsf{Ext}(\tilde{w}; x'_R)$
- $R' = z'_R \oplus y'_R$
- $L' = f_7(x_{live}, y'_1, x_R, y'_2, z_R, y'_3, L)$
- $k' = \mathsf{Dec}(L', R')$
- $t = \mathsf{Tag}_{k'}(m)$
- $(m', t') = f_8(x_{live}, y'_1, x_R, y'_2, z_R, y'_3, L, m, t)$
- Output $x_{live}, y'_{live}, x_R, z_R, L, m, t$

$View2^m_{f_1,..,f_8}$:

- $w \leftarrow W$
- $x_{live} \in_R \{0,1\}^d$, $z_R \in_R \{0,1\}^l$, $k \in_R \{0,1\}^d, x_R \in_R \{0,1\}^d$
- $x'_{live} = f_1(x_{live})$
- $y_1||y_2||y_3 = y_{live} = \mathsf{Ext}'(w; x_{live})$
- $y'_1||y'_2||y'_3 = y'_{live} = \mathsf{Ext}'(w; x'_{live})$
- $x'_R = f_3(x_{live}, y'_1, x_R)$
- $(L, R) \leftarrow \mathsf{Enc}(k)$
- $y_R = z_R \oplus R$
- $z'_R = f_5(x_{live}, y'_1, x_R, y'_2, z_R)$
- If $\nexists \tilde{w} \in \mathsf{Support}(\tilde{W})$ such that $\mathsf{Ext}(\tilde{W}; x_R) = y_R$, then set $\tilde{w} = \perp$ and Output $\perp$
  else $\tilde{w} \leftarrow \tilde{W}|(\mathsf{Ext}(\tilde{W}; x_R) = y_R)$
  $\tilde{W}$ is the conditional source, $\tilde{W} := W|(\mathsf{Ext}'(W; x_{live}) = y_{live}, \mathsf{Ext}'(W; x'_{live}) = y'_{live})$
- $y'_R = \mathsf{Ext}(\tilde{w}; x'_R)$
- $R' = z'_R \oplus y'_R$
- $L' = f_7(x_{live}, y'_1, x_R, y'_2, z_R, y'_3, L)$
- $k' = \mathsf{Dec}(L', R')$
- $t = \mathsf{Tag}_{k'}(m)$
- $(m', t') = f_8(x_{live}, y'_1, x_R, y'_2, z_R, y'_3, L, m, t)$
- Output $x_{live}, y'_{live}, x_R, z_R, L, m, t$

<table>
<tr>
<td>

$\underline{View3^m_{f_1,..,f_8}}$:

- $w \leftarrow W$
- $x_{live} \in_R \{0,1\}^d, z_R \in_R \{0,1\}^l$, $k \in_R \{0,1\}^d, x_R \in_R \{0,1\}^d$
- $x'_{live} = f_1(x_{live})$
- $y_1||y_2||y_3 = y_{live} = \mathsf{Ext}'(w; x_{live})$
- $y'_1||y'_2||y'_3 = y'_{live} = \mathsf{Ext}'(w; x'_{live})$
- $L, k' \leftarrow Tamper^k_{f,g}$ *f, g will be hardwired with $x_{live}, y_{live}, y'_{live}, x_R, z_R$
- $t = \mathsf{Tag}_{k'}(m)$
- $(m', t') = f_8(x_{live}, y'_1, x_R, y'_2, z_R, y'_3, L, m, t)$
- Output $x_{live}, y'_{live}, x_R, z_R, L, m, t$

* Description of f, g is given in Claim 6

</td>
<td>

$\underline{View4^m_{f_1,..,f_8}}$:

- $w \leftarrow W$
- $x_{live} \in_R \{0,1\}^d, z_R \in_R \{0,1\}^l$, $k \in_R \{0,1\}^d, x_R \in_R \{0,1\}^d$
- $x'_{live} = f_1(x_{live})$
- $y_1||y_2||y_3 = y_{live} = \mathsf{Ext}'(w; x_{live})$
- $y'_1||y'_2||y'_3 = y'_{live} = \mathsf{Ext}'(w; x'_{live})$
- $L, k' = Copy(k, Sim_{f,g})$ *f, g will be hardwired with $x_{live}, y_{live}, y'_{live}, x_R, z_R$
- $t = \mathsf{Tag}_{k'}(m)$
- $(m', t') = f_8(x_{live}, y'_1, x_R, y'_2, z_R, y'_3, L, m, t)$
- Output $x_{live}, y'_{live}, x_R, z_R, L, m, t$

* Description of f, g is given in Claim 6

</td>
</tr>
</table>

We define the random variables corresponding to the views described above as follows.

- $View1^m_{f_1,..,f_8} \equiv (X_{live}, Y'_{live}, X_R, Z^1_R, L, m, T^1)$
- $View2^m_{f_1,..,f_8} \equiv (X_{live}, Y'_{live}, X_R, Z^2_R, L, m, T^2)$
- $View3^m_{f_1,..,f_8} \equiv (X_{live}, Y'_{live}, X_R, Z^3_R, L, m, T^3)$
- $View4^m_{f_1,..,f_8} \equiv (X_{live}, Y'_{live}, X_R, Z^3_R, L^4, m, T^4)$

In the above description, we superscript a random variable with the corresponding view number i, wherever there is a change in distribution from the previous view.

We consider the following claims to complete the hybrid argument and then bound the success probability of Eve given the final view ($View4^m_{f_1,..,f_8}$) to complete the proof.

Moving from $View0^m_{f_1,\cdots,f_8}$ to $View1^m_{f_1,\cdots,f_8}$: In the first hybrid, we wish to analyze Eve's success probability, given an identical view, where we use a conditional source $\tilde{W} = W|(\mathsf{Ext}'(W; x_{live}) = y_{live}, \mathsf{Ext}'(W; x'_{live}) = y'_{live})$ (post drawing the liveness test responses and seed from the same distribution as in $View0^m_{f_1,\cdots,f_8}$ and then fixing them) for further extractions. The use of a different sample of w for the liveness test is crucial and the reason for doing this becomes clear when we move to $View3^m_{f_1,\cdots,f_8}$. We show how the two views are identical and then show why Eve's success probability remains the same.

Claim 4

$$\Pr[(m', t') \leftarrow \mathsf{Eve}(View0^m_{f_1,\cdots,f_8}) \wedge m \neq m' \wedge \mathsf{Vrfy}_K(m', t') = 1]$$
$$= \Pr[(m', t') \leftarrow \mathsf{Eve}(View1^m_{f_1,\cdots,f_8}) \wedge m \neq m' \wedge \mathsf{Vrfy}_K(m', t') = 1]$$

where the probabilities are taken over the randomness used to generate $View0^m_{f_1,\cdots,f_8}$ *and* $View1^m_{f_1,\cdots,f_8}$ *respectively, i.e.,* W, X_{live}, X_R, K*, the randomness used in* Enc *and* $W, \tilde{W}, X_{live}, X_R, K$*, the randomness used in* Enc *respectively.*

Proof. Taking $A = W$, $B = X_{live}$ and $C = (Y_{live} = \mathsf{Ext}'(W; X_{live}), Y'_{live} = \mathsf{Ext}'(W; X'_{live})) = f(A, B)$ in Lemma 1, we get:

$$W, X_{live}, Y_{live}, Y'_{live} \equiv \tilde{W}, X_{live}, Y_{live}, Y'_{live}$$

where $\tilde{W} \equiv W | \mathsf{Ext}'(W; X_{live}) = Y_{live}, \mathsf{Ext}'(W; X'_{live}) = Y'_{live}$. Further, as K and X_R is independent of the random variables above, we get:

$$K, X_R, W, X_{live}, Y_{live}, Y'_{live} \equiv K, X_R, \tilde{W}, X_{live}, Y_{live}, Y'_{live}$$

The randomness used in Enc is independent of the random variables above. Hence, Z_R, L and T can be obtained as functions of the above random variables (and the randomness used in Enc). Then, by using Lemma 3, we get:

$$K, X_{live}, Y'_{live}, X_R, Z_R, L, m, T \equiv K, X_{live}, Y'_{live}, X_R, Z^1_R, L, m, T^1$$
$$K, View0^m_{f_1,\cdots,f_8} \equiv K, View1^m_{f_1,\cdots,f_8}$$

Again as Eve's output and the verification check are a function of the above random variables, by use of Lemma 3, it follows that

$$\Pr[(m', t') \leftarrow \mathsf{Eve}(View0^m_{f_1,\cdots,f_8}) \wedge m \neq m' \wedge \mathsf{Vrfy}_k(m', t') = 1]$$
$$= \Pr[(m', t') \leftarrow \mathsf{Eve}(View1^m_{f_1,\cdots,f_8}) \wedge m \neq m' \wedge \mathsf{Vrfy}_k(m', t') = 1]$$

□

Moving from $View1^m_{f_1,\cdots,f_8}$ to $View2^m_{f_1,\cdots,f_8}$: We replace Z_R with U and then sample the source consistently (upto some error) in $View2^m_{f_1,\cdots,f_8}$. This reverse sampling of the source becomes a little complicated as it has to be not only consistent with the z_R sampled but also has to be consistent with the liveness test responses. This is why we would consistently reverse sample from the conditional source $\tilde{W}$ here. Now, showing that Eve's success probability in $View1^m_{f_1,\cdots,f_8}$ is at most her success probability in this view (upto some error) captures that R remains hidden from Eve.

Claim 5 *If* Ext *is an* (n, t, d, l, ϵ_2)*- average case extractor, then*

$$\Pr[(m', t') \leftarrow \mathsf{Eve}(View1^m_{f_1,\cdots,f_8}) \wedge m \neq m' \wedge \mathsf{Vrfy}_K(m', t') = 1]$$
$$\leq \Pr[(m', t') \leftarrow \mathsf{Eve}(View2^m_{f_1,\cdots,f_8}) \wedge m \neq m' \wedge \mathsf{Vrfy}_K(m', t') = 1] + 2^{-\lambda} + 2\epsilon_1$$

where the probabilities are taken over the randomness used to generate $View1^m_{f_1,\cdots,f_8}$ *and* $View2^m_{f_1,\cdots,f_8}$ *respectively.*

Proof. In order to use the extractor security, we first need to ensure that $\tilde{W}$ has "high enough entropy". We define the following good set:

$$\mathcal{G} = \{(x_{live}, y_{live}, y'_{live}) : \mathbf{H}_\infty(\tilde{W}) = \\ \mathbf{H}_\infty(W|\mathsf{Ext}'(W; x_{live}) = y_{live}, \mathsf{Ext}'(W; x'_{live}) = y'_{live}) \geq t' - 6l' - \lambda\}$$

We now define the good event:

$$\mathsf{Good} : (X_{live}, Y_{live}, Y'_{live}) \in \mathcal{G}$$

Here, Y_{live} and Y'_{live} denote the random variables $Y_{live} = \mathsf{Ext}'(W; X_{live})$ and $Y'_{live} = \mathsf{Ext}'(W; X'_{live})$. Let Good^C denote its complement event. Consider, by Proposition 1(Appendix A)

$$\begin{aligned}&\mathbf{SD}\left((K, View1^m_{f_1,\cdots,f_8}); (K, View2^m_{f_1,\cdots,f_8})\right) \\ &\leq \mathbf{SD}_{\mathsf{Good}}((K, View1^m_{f_1,\cdots,f_8}); (K, View2^m_{f_1,\cdots,f_8})). \Pr[\mathsf{Good}] \\ &+ \mathbf{SD}_{\mathsf{Good}^C}((K, View1^m_{f_1,\cdots,f_8}); (K, View2^m_{f_1,\cdots,f_8})). \Pr[\mathsf{Good}^C] \end{aligned} \tag{10}$$

where the subscript notation is used to denote the statistical distance conditioned on the specific event (in the subscript). By Lemma 4, we get:

$$\begin{aligned}\Pr[\mathsf{Good}^C] &= \Pr[(X_{live}, Y_{live}, Y'_{live}) \notin \mathcal{G}] \\ &= \Pr_{y_{live}, y'_{live}}[\mathbf{H}_\infty(W|Y_{live} = y_{live}, Y'_{live} = y'_{live}) < t' - 6l' - \lambda] \\ &\leq \Pr_{y_{live}, y'_{live}}[\mathbf{H}_\infty(W|Y_{live} = y_{live}, Y'_{live} = y'_{live}) < \widetilde{\mathbf{H}}_\infty(W|Y_{live}, Y'_{live}) - \lambda] \\ &\leq 2^{-\lambda}\end{aligned} \tag{11}$$

Let W_1 denote the random variable:

- If $\nexists \tilde{w} \in \text{Support}(\tilde{W})$ such that $\mathsf{Ext}(\tilde{W}; x_R) = y_R$, then set $\tilde{w} = \perp$ and Output $\perp$
- else $\tilde{w} \leftarrow \tilde{W}|\mathsf{Ext}(\tilde{W}; x_R) = y_R$

By setting parameters appropriately, we ensure $\mathbf{H}_\infty(\tilde{W}) \geq t$, where t is the min entropy required for using $\mathsf{Ext}(.)$. Then, by Lemma 5, we know that

$$\forall (x_{live}, y_{live}, y'_{live}) \in \mathcal{G},\ \tilde{W}, X_R, Y^1_R \approx_{2\epsilon_1} W_1, X_R, Y^2_R$$

where the distributions which change in the two views have been superscripted with the corresponding view number. Then, by using Proposition 2(Appendix A), with $A \equiv ((\tilde{W}, X_R, Y^1_R)|\mathsf{Good})$, $B \equiv ((W_1, X_R, Y^2_R)|\mathsf{Good})$ and $C \equiv ((X_{live}, Y_{live}, Y'_{live})|\mathsf{Good})$, we get:

$$X_{live}, Y_{live}, Y'_{live}, \tilde{W}, X_R, Y^1_R|\mathsf{Good} \approx_{2\epsilon_1} X_{live}, Y_{live}, Y'_{live}, W_1, X_R, Y^2_R|\mathsf{Good}$$
$$X_{live}, Y'_{live}, \tilde{W}, X_R, Y^1_R|\mathsf{Good} \approx_{2\epsilon_1} X_{live}, Y'_{live}, W_1, X_R, Y^2_R|\mathsf{Good}$$

Further, as K is independent of the above random variables, we get:

$$K, X_{live}, Y'_{live}, \tilde{W}, X_R, Y_R^1 | \mathsf{Good} \approx_{2\epsilon_1} K, X_{live}, Y'_{live}, W_1, X_R, Y_R^2 | \mathsf{Good}$$

The randomness used in Enc is independent of the above random variables. Hence, Z_R^1, L and T^1 can be obtained as a function of the above random variables (and the randomness used in Enc). Then, by using Lemma 3, we get:

$$K, X_{live}, Y'_{live}, X_R, Z_R^1, L, m, T^1 | \mathsf{Good} \approx_{2\epsilon_1} K, X_{live}, Y'_{live}, X_R, Z_R^2, L, m, T^2 | \mathsf{Good}$$

$$K, View1^m_{f_1,\cdots,f_8} | \mathsf{Good} \approx_{2\epsilon_1} K, View2^m_{f_1,\cdots,f_8} | \mathsf{Good} \quad (12)$$

Then, by using Equations 11 and 12 in Equation 10, we get:

$$\mathbf{SD}\left((K, View1^m_{f_1,\cdots,f_8}); (K, View2^m_{f_1,\cdots,f_8})\right) \leq 2^{-\lambda} + 2\epsilon_1$$

Finally, by use of Lemma 3, we get the desired bound:

$$\begin{aligned} &\Pr[(m',t') \leftarrow \mathsf{Eve}(View1^m_{f_1,\cdots,f_8}) \wedge m \neq m' \wedge \mathsf{Vrfy}_K(m',t') = 1] \\ &\leq \Pr[(m',t') \leftarrow \mathsf{Eve}(View2^m_{f_1,\cdots,f_8}) \wedge m \neq m' \wedge \mathsf{Vrfy}_K(m',t') = 1] + 2^{-\lambda} + 2\epsilon_1 \end{aligned}$$

□

Moving from $View2^m_{f_1,\cdots,f_8}$ to $View3^m_{f_1,\cdots,f_8}$: In $View3^m_{f_1,\cdots,f_8}$, we want to capture the tampering on k by the tamper random variable of the augmented NMC, $Tamper^k_{f,g}$. To be able to to do this, we have to first capture the tampering on L and R as a correct split-state tampering by (f,g). In order to describe the functions, we would need to hardwire the liveness test seed and responses, x_R and z_R. Now, to get the tampering of R, a w consistent with the hardwired values has to be sampled. But this sampler might return $\perp$. However, as the function g cannot output $\perp$, we replace $\tilde{w}$ with an arbitrary string, whenever the sampling returns $\perp$. We now analyze Eve's success probability in this modified view.

Claim 6

$$\begin{aligned} &\Pr[(m',t') \leftarrow \mathsf{Eve}(View2^m_{f_1,\cdots,f_8}) \wedge m \neq m' \wedge \mathsf{Vrfy}_K(m',t') = 1] \\ &\leq \Pr[(m',t') \leftarrow \mathsf{Eve}(View3^m_{f_1,\cdots,f_8}) \wedge m \neq m' \wedge \mathsf{Vrfy}_K(m',t') = 1] \end{aligned}$$

where the probabilities are taken over the randomness used to generate $View2^m_{f_1,\cdots,f_8}$ and $View3^m_{f_1,\cdots,f_8}$ respectively.

Proof. We define the tampering functions f, g hardwired with $x_{live}, y_{live}, y'_{live}, x_R, z_R$ as follows.

$f_{x_{live},y_{live},y'_{live},x_R,z_R}(L)$:

- Output $L' = f_7(x_{live}, y'_1, x_R, y'_2, z_R, y'_3, L)$

$g_{x_{live},y_{live},y'_{live},x_R,z_R}(R)$:

- $\tilde{w} \leftarrow \tilde{W} | \mathsf{Ext}(\tilde{W}; x_R) = z_R \oplus R$
- If $\tilde{w} = \perp$, set $\tilde{w} := 0$.
- $y'_R = \mathsf{Ext}(\tilde{w}; f_1(x_R))$
- $z'_R = f_5(x_{live}, y'_1, x_R, y'_2, z_R)$
- Output $R' = z'_R \oplus y'_R$

The function g is a randomized function here (atypical to tampering function descriptions). But, the randomness required for this sampling can be sampled a priori and hardwired in g, along with the other values, making it a deterministic function. Hence, while we use the above description of g for simplicity, it is simple to convert it to a deterministic function. For the sake of simplicity we avoid explicitly writing the hardwired values while referring to the tampering functions.
Let W_1 be the following distribution

- If $\nexists \tilde{w} \in \text{Support}(\tilde{W})$ such that $\mathsf{Ext}(\tilde{W}; x_R) = y_R$, then set $\tilde{w} = \perp$ and Output $\perp$
- else $\tilde{w} \leftarrow \tilde{W} | \mathsf{Ext}(\tilde{W}; x_R) = y_R$

Observe that

$$(View3^m_{f_1,\cdots,f_8} | W_1 \neq \perp) \equiv (View2^m_{f_1,\cdots,f_8} | W_1 \neq \perp)$$

Hence, as K is independent of the event $W_1 \neq \perp$, $((K, View3^m_{f_1,\cdots,f_8}) | W_1 \neq \perp) \equiv (K, (View3^m_{f_1,\cdots,f_8} | W_1 \neq \perp)) \equiv (K, (View2^m_{f_1,\cdots,f_8} | W_1 \neq \perp)) \equiv ((K, View2^m_{f_1,\cdots,f_8}) | W_1 \neq \perp)$. Hence, Proposition 1 (Appendix A), we get:

$$\begin{aligned}
&\mathbf{SD}\left((K, View2^m_{f_1,\cdots,f_8}); (K, View3^m_{f_1,\cdots,f_8})\right) \\
&\leq \mathbf{SD}\left((K, View2^m_{f_1,\cdots,f_8}); (K, View3^m_{f_1,\cdots,f_8}) | W_1 \neq \perp\right) + \Pr[W_1 = \perp] \\
&= 0 + \Pr[W_1 = \perp] \\
&\leq \epsilon_2
\end{aligned}$$

The last inequality follows from Inequality 2 in proof of Lemma 5 where W^{res} is W_1. □

<u>**Moving from $View3^m_{f_1,\cdots,f_8}$ to $View4^m_{f_1,\cdots,f_8}$:**</u> To use MAC security, it is crucial that we argue the non-malleability of the MAC key. For this, we use the non-malleability of $(\mathsf{Enc}, \mathsf{Dec})$. To do this, we observe that tampering functions f, g are indeed split-state, as the hardwired values $(x_{live}, y_{live}, y'_{live}, x_R, z_R)$ are independent of the two states L and R.[5] Then we analyze Eve's success probability in this modified view.

Claim 7 *If* $(\mathsf{Enc}, \mathsf{Dec})$ *is an* ϵ_3*- augmented non-malleable code, then*

$$\begin{aligned}
&\Pr[(m', t') \leftarrow \mathsf{Eve}(View3^m_{f_1,\cdots,f_8}) \wedge m \neq m' \wedge \mathsf{Vrfy}_K(m', t') = 1] \\
&\leq \Pr[(m', t') \leftarrow \mathsf{Eve}(View4^m_{f_1,\cdots,f_8}) \wedge m \neq m' \wedge \mathsf{Vrfy}_K(m', t') = 1] + \epsilon_3
\end{aligned}$$

where the probabilities are taken over the randomness used to generate $View3^m_{f_1,\cdots,f_8}$ *and* $View4^m_{f_1,\cdots,f_8}$ *respectively.*

[5] As mentioned while describing g in $View3^m_{f_1,\cdots,f_8}$, although the given description of g is randomized, but by fixing the randomness it can be made deterministic.

Proof. As already mentioned the tampering functions are split-state. Hence, by the security of $(\mathsf{Enc}, \mathsf{Dec})$ we have

$$\forall(x_{live}, y_{live}, y'_{live}, x_R, z_R), \quad \forall k, \quad Tamper^k_{f,g} \approx_{\epsilon_3} Copy(k, Sim_{f,g})$$

where the message to be encoded is k and the split-state tampering functions are hardwired with $(x_{live}, y_{live}, y'_{live}, x_R, z_R)$. Hence, by Proposition 2(Appendix A) with A, B, C being $Tamper^K_{f,g},\ Copy(K, Sim_{f,g}),\ (K, X_{live}, Y_{live}, Y'_{live}, X_R, Z^3_R)$ respectively, we have

$$\begin{aligned}&K, X_{live}, Y_{live}, Y'_{live}, X_R, Z^3_R, Tamper^K_{f,g}\\ &\approx_{\epsilon_3} K, X_{live}, Y_{live}, Y'_{live}, X_R, Z^3_R, Copy(K, Sim_{f,g})\end{aligned}$$

For clarity, we denote the random variables (L, K', T) of $View3$ and $View4$ by (L, K'^3, T^3) and (L^4, K'^4, T^4) respectively.

$$K, X_{live}, Y_{live}, Y'_{live}, X_R, Z^3_R, L, K'^3 \approx_{\epsilon_3} K, X_{live}, Y_{live}, Y'_{live}, X_R, Z^3_R, L^4, K'^4$$
$$K, X_{live}, Y_{live}, Y'_{live}, X_R, Z^3_R, L, m, T^3 \approx_{\epsilon_3} K, X_{live}, Y_{live}, Y'_{live}, X_R, Z^3_R, L^4, m, T^4$$
$$K, View3^m_{f_1,\cdots,f_8} \approx_{\epsilon_3} K, View4^m_{f_1,\cdots,f_8}$$

Above implications follow from Lemma 3. Therefore

$$\begin{aligned}&\Pr[(m', t') \leftarrow \mathsf{Eve}(View3^m_{f_1,\cdots,f_8}) \wedge m \neq m' \wedge \mathsf{Vrfy}_K(m', t') = 1]\\ &\leq \Pr[(m', t') \leftarrow \mathsf{Eve}(View4^m_{f_1,\cdots,f_8}) \wedge m \neq m' \wedge \mathsf{Vrfy}_K(m', t') = 1] + \epsilon_3\end{aligned}$$

□

We now combine the above claims with MAC security and show how to get the desired bound on Eve's success probability in the synchronous case.

Claim 8 *If* $(\mathsf{Tag}, \mathsf{Vrfy})$ *is an* ϵ_4*- one time MAC (the auxiliary information variant defined in Section 3) then*

$$\Pr[(m', t') \leftarrow \mathsf{Eve}(View0^m_{f_1,\cdots,f_8}) \wedge m \neq m' \wedge \mathsf{Vrfy}_K(m', t') = 1] \leq 2^{-\lambda} + 3\epsilon_2 + \epsilon_3 + \epsilon_4$$

where the probability is taken over the randomness used to generate $View0^m_{f_1,\cdots,f_8}$ *respectively.*

Proof. Combining Claims 4,5,6,7 we get

$$\begin{aligned}&\Pr[(m', t') \leftarrow \mathsf{Eve}(View0^m_{f_1,\cdots,f_8}) \wedge m \neq m' \wedge \mathsf{Vrfy}_K(m', t') = 1]\\ &\leq \Pr[(m', t') \leftarrow \mathsf{Eve}(View4^m_{f_1,\cdots,f_8}) \wedge m \neq m' \wedge \mathsf{Vrfy}_K(m', t') = 1] + 2^{-\lambda} + 3\epsilon_2 + \epsilon_3\end{aligned} \tag{13}$$

We now consider the following events with respect to $View4^m_{f_1,\cdots,f_8}$.

<u>$Case1$:</u> $Sim_{f,g}$ does not output $same^*$

$$K \equiv U_\tau$$

$$K, X_{live}, Y_{live}, Y'_{live}, X_R, Z_R^3, Sim_{f,g}|Case1 \equiv U_\tau, X_{live}, Y_{live}, Y'_{live}, X_R, Z_R^3, Sim_{f,g}|Case1 \quad (14)$$

$$K, m, X_{live}, Y'_{live}, X_R, Z_R^3, L^4, K'^4 \equiv U_\tau, m, X_{live}, Y'_{live}, X_R, Z_R^3, L^4, K'^4$$

$$K, m, X_{live}, Y'_{live}, X_R, Z_R^3, L^4, \mathsf{Tag}_{K'^4}(m) \equiv U_\tau, m, X_{live}, Y'_{live}, X_R, Z_R^3, L^4, \mathsf{Tag}_{K'^4}(m)$$

Implication 14 follows from Proposition 5(Appendix A) with A, B, C, E being $K, (X_{live}, Y_{live}, Y'_{live}, X_R, Z_R^3, Sim_{f,g}), U_\tau, Case1$ respectively. Therefore, given $Case1$, Eve's view is $View4^m|Case1 \equiv (m, X_{live}, Y'_{live}, X_R, Z_R^3, L^4, \mathsf{Tag}_{K'^4}(m))$ is independent of MAC key K. This is because the randomness used to generate $View4^m|Case1$, which is $X_{live}, X_R, W, \tilde{W}$, the randomness used in $Sim_{f,g}$, are all independent of K. Then, as $E \equiv View4^m|Case1$ is independent of K, by the MAC security we get:

$$\begin{aligned}
&\Pr[(m', t') \leftarrow \mathsf{Eve}(View4^m) \wedge m \neq m' \wedge \mathsf{Vrfy}_K(m', t') = 1|Case1] \\
&= \Pr_k[\mathsf{Tag}_k(m') = t' \wedge (m \neq m')|E = (m, x_{live}, y'_{live}, x_R, z_R, L, Tag_{k'}(m))] \\
&\leq \epsilon_4
\end{aligned} \quad (15)$$

<u>Case2:</u> $Sim_{f,g}$ outputs $same^*$

$$K \equiv U_\tau$$

$$K, X_{live}, Y_{live}, Y'_{live}, X_R, Z_R^3, Sim_{f,g}|Case2 \equiv U_\tau, X_{live}, Y_{live}, Y'_{live}, X_R, Z_R^3, Sim_{f,g}|Case2 \quad (16)$$

$$K, m, X_{live}, Y'_{live}, X_R, Z_R^3, L^4|Case2 \equiv U_\tau, m, X_{live}, Y'_{live}, X_R, Z_R^3, L^4|Case2$$

Implication 16 follows from Proposition 5(Appendix A) with A, B, C, E being $K, (X_{live}, Y_{live}, Y'_{live}, X_R, Z_R^3, Sim_{f,g}), U_\tau$, Case2 respectively. Therefore, given Case2, Eve's view is $View4^m|Case2 \equiv (m, X_{live}, Y'_{live}, X_R, Z_R^3, L^4, \mathsf{Tag}_K(m))$. The only information Eve has regarding K is $\mathsf{Tag}_K(m)$. Then, as $E \equiv (m, X_{live}, Y'_{live}, X_R, Z_R^3, L^4)|Case2$ is independent of K, by MAC security we get:

$$\begin{aligned}
&\Pr[(m', t') \leftarrow \mathsf{Eve}(View4^m) \wedge m \neq m' \wedge \mathsf{Vrfy}_K(m', t') = 1|Case2] \\
&= \Pr_k[\mathsf{Tag}_k(m') = t' \wedge (m' \neq m)|\mathsf{Tag}_k(m) = t, E = (m, x_{live}, y'_{live}, x_R, z_R, L)] \\
&= \Pr_k[\mathsf{Tag}_k(m') = t' \wedge (m' \neq m)|\mathsf{Tag}_k(m) = t] \\
&\leq \epsilon_4
\end{aligned} \quad (17)$$

Combining inequalities 15,17 with the inequality 13 gives

$$\Pr[(m', t') \leftarrow \mathsf{Eve}(View0^m) \wedge m \neq m' \wedge \mathsf{Vrfy}_K(m', t') = 1] \leq 2^{-\lambda} + 3\epsilon_2 + \epsilon_3 + \epsilon_4$$

□

Using Claim 8 and Equation 9, we get:

$$\Pr[\mathsf{Eve\ wins}|\mathsf{Sync}, \mathsf{Pass}] \leq 2^{-\lambda} + 3\epsilon_2 + \epsilon_3 + \epsilon_4$$

Hence, Lemma 8 is proved. □

Now, combining Lemmata 7 and 8, Equation 4 gives:

$$\Pr[\mathsf{Eve}\text{ wins}] =\leq 2.(2^{-l'} + 2^{3l'}\epsilon_1 + 2^{-\lambda} + 2^{-\lambda} + 3\epsilon_2 + \epsilon_3 + \epsilon_4)$$

We set κ such that $2^{-\kappa} = 2.(2^{-l'} + 2^{3l'}\epsilon_1 + 2.2^{-\lambda} + 3\epsilon_2 + \epsilon_3 + \epsilon_4)$. Thus robustness of message authentication protocol is proved.

5.2 Proof of Theorem 2

We now prove that π^{PA} is a Privacy Amplification protocol.
Correctness The correctness of π^{PA} follows by the correctness of π^{AUTH}.
Robustness: We need to show

$$\Pr[K_A \neq K_B \wedge K_A \neq \perp \wedge K_B \neq \perp] \leq 2^{-\kappa}$$

$\Pr[K_A \neq K_B \wedge K_A \neq \perp \wedge K_B \neq \perp]$

$$= \Pr[M_A \neq M_B \wedge M_A \neq \perp \wedge M_B \neq \perp]. \Pr[K_A \neq K_B | M_A \neq M_B \wedge M_A \neq \perp \wedge M_B \neq \perp]$$
$$\leq 2^{-\kappa} \text{ (by robustness of } \pi^{\mathsf{AUTH}})$$

Extraction: $Sent_a, Sent_b$ denote the messages sent by Alice and Bob upon execution of π^{PA} in presence of Eve, the pair $Sent = (Sent_a, Sent_b)$ contains an active Eve's view of the protocol. For extraction we need to show

- If $K_B \neq \perp$, then $K_B, Sent \approx_{\epsilon_5} U_{l''}, Sent$
- If $K_A \neq \perp$, then $K_A, Sent \approx_{\epsilon_5} U_{l''}, Sent$

If $K_B \neq \perp$, K_B is the extractor output on an independent uniform seed $M_B \neq \perp$. As M_B is independent of $X_{live}, Y_{live}, Y'_{live}, X_R, Z_R, L, K, K'$, by the use of average case extractors we have,

$$K_B, M_B, X_{live}, Y_{live}, Y'_{live}, X_R, Z_R, L, K, K' \approx_{\epsilon_5} U_{l''}, M_B, X_{live}, Y_{live}, Y'_{live}, X_R, Z_R, L, K, K'$$
$$K_B, M_B, X_{live}, Y_{live}, Y'_{live}, X_R, Z_R, L, K, K', T \approx_{\epsilon_5} U_{l''}, M_B, X_{live}, Y_{live}, Y'_{live}, X_R, Z_R, L, K, K', T$$
$$K_B, Sent \approx_{\epsilon_5} U_{l''}, Sent$$

$K_A \neq \perp \Rightarrow M_A \neq \perp \wedge M_B \neq \perp$.
We can write, $\mathbf{SD}\left((K_A, Sent); (U_{l''}, Sent)\right)$

$$= \Pr[M_A = M_B \wedge M_A \neq \perp \wedge M_B \neq \perp]\mathbf{SD}_{M_A = M_B}((K_A, Sent); (U_{l''}, Sent))$$
$$+ \Pr[M_A \neq M_B \wedge M_A \neq \perp \wedge M_B \neq \perp]\mathbf{SD}_{M_A \neq M_B}((K_A, Sent); (U_{l''}, Sent))$$
$$\leq \mathbf{SD}\left((K_B, Sent); (U_{l''}, Sent)\right) + \Pr[M_A \neq M_B \wedge M_A \neq \perp \wedge M_B \neq \perp]$$
$$\leq \epsilon_5 + 2^{-\kappa}$$

5.3 Analysis of Entropy loss and Other Parameters

To get desired parameters as in Theorem 2, we use optimal constructions of building blocks given in following theorems.

Lemma 9. *[GUV07] For every constant $\nu > 0$ all integers $n \geq t$ and all $\epsilon \geq 0$, there is an explicit (efficient) $(n, t, d, l, \epsilon)-$strong extractor with $l = (1 - \nu)t - \mathcal{O}(\log(n) + \log(\frac{1}{\epsilon}))$ and $d = \mathcal{O}(\log(n) + \log(\frac{1}{\epsilon}))$.*

Now, as we give some auxiliary information about the source, we require the security of the extractor to hold, even given this information. Hence, we use average case extractors, given in the following lemma.

Lemma 10. *[DORS08] For any $\mu > 0$, if* Ext *is a (worst case)$(n, t, d, l, \epsilon)-$strong extractor, then* Ext *is also an average-case $(n, t + \log(\frac{1}{\mu}), d, l, \epsilon + \mu)$ strong extractor.*

Now, we also encode the authentication keys and tags using the underlying non-malleable code. Hence, we require them to have short lengths. This is guaranteed by the following lemma [JKS93]:

Lemma 11. *For any $n', \varepsilon_2 > 0$ there is an efficient ε_2-secure one time* MAC *with $\delta \leq (\log(n') + \log(\frac{1}{\varepsilon_2}))$, $\tau \leq 2\delta$, where τ, n', δ are key, message, tag length respectively.*

We now set the parameters:

- For the MAC, we set:
  - $\epsilon_4 = 2^{-\lambda}$
  - Key length: $\tau = \mathcal{O}(\lambda)$
  - Tag length: $\delta = \mathcal{O}(\tau) = \mathcal{O}(\lambda)$
  - Message length: d, will be set below.
- For the liveness test Extractor, we set:
  - $\epsilon_1 = 2^{-4.\lambda}$
  - Seed length: $d = \mathcal{O}(\log n + 4.\lambda)$
  - output length: $3l' = \mathcal{O}(\lambda)$
- Now, we calculate the entropy loss:
  - From the transcript of the protocol, the entropy loss that occurs is: $3l' + l + \delta = 3l' + l + \mathcal{O}(\lambda)$
  - Additional leakage results in a loss: $\mathcal{O}(\lambda) + 3l'$
  - Hence, we require $\mathbf{H}_\infty(W) - (6l' + l + \mathcal{O}(\lambda)) \geq max\{t, t', t''\}$
  - Then, by setting $\epsilon_5 = 2^{-\lambda+1}$, we know $l'' = (1 - \mu)t'' - \mathcal{O}(\log n + \lambda)$, we get a total entropy loss $= 6l' + l + \mathcal{O}(\lambda) + \mathcal{O}(\log n + \lambda) = \mathcal{O}(\lambda) + l + \mathcal{O}(\log n + \lambda) = l + \mathcal{O}(\log n + \lambda)$

To finally evaluate the entropy loss, we set parameters for the NMC:

2A: If we consider a constant rate optimal error NMC, we set:
  - We know message length: $\tau = \mathcal{O}(\lambda)$
  - $\epsilon_3 = 2^{-\mathcal{O}(\lambda)}$
  - Codeword length: $2l = \mathcal{O}(\lambda)$

  Then entropy loss $= (l + \mathcal{O}(\log n + \lambda)) = \mathcal{O}(\lambda) + \mathcal{O}(\log n + \lambda) = \mathcal{O}(\log n + \lambda)$

2B: If we instantiate our construction using the NMC [Li17], we set:
- We know message length: $\tau = \mathcal{O}(\lambda)$
- $\epsilon_3 = 2^{-\mathcal{O}(\lambda)}$
- Codeword length: $2l = \mathcal{O}(\lambda \log \lambda)$

Then entropy loss $= (l+\mathcal{O}(\log n+\lambda)) = \mathcal{O}(\lambda \log \lambda)+\mathcal{O}(\log n+\lambda) = \mathcal{O}(\log n+\lambda \log \lambda)$

– Finally, as we set $2^{-\kappa} == 2.(2^{-l'} + 2^{3l'}\epsilon_1 + 2.2^{-\lambda} + 3\epsilon_2 + \epsilon_3 + \epsilon_4)$. By setting $\epsilon_2 = 2^{-\lambda}$, in both 2A and 2B, we get $2^{-\kappa} = 2^{-\mathcal{O}(\lambda)}$. Hence, $\kappa = \mathcal{O}(\lambda)$.
– The error in Extraction property of $\pi^{\mathsf{PA}} = \epsilon_5 + 2^{-\kappa} = 2^{-\lambda} + 2^{-\kappa} = 2^{-\kappa+1}$

5.4 Privacy Amplification from Augmented-NMREs

Theorem 3 *If* $(\mathsf{Enc}, \mathsf{Dec})$ *in* π^{AUTH} *is a two-state, constant rate, optimal error augmented non-malleable randomness encoder, then the 8-round protocol* π^{PA} *in Figure 1 is a* $(t', l'', \kappa, \kappa+1)$*-secure privacy amplification protocol with optimal entropy loss with min-entropy requirement* $\Omega(\log(n) + \kappa)$*.*

Proof: The only modification made in this protocol is that instead of picking the MAC key k uniformly at random and then encoding it using NMCs, we use the key k and its encoding output by the NMRE. As augmented-NMREs guarantee the K looks uniform even given L and the modified key K', the proof structure of this theorem follows on the same lines as the security proof in Sections 5.1 and 5.2.

6 Acknowledgement

The second author thanks Yevgeniy Dodis and Leonid Reyzin for several insightful discussions.

References

AAG+16. Divesh Aggarwal, Shashank Agrawal, Divya Gupta, Hemanta K. Maji, Omkant Pandey, and Manoj Prabhakaran. Optimal computational split-state non-malleable codes. In *Theory of Cryptography - 13th International Conference, TCC 2016-A, Tel Aviv, Israel, January 10-13, 2016, Proceedings, Part II*, pages 393–417, 2016.

ADKO15. Divesh Aggarwal, Yevgeniy Dodis, Tomasz Kazana, and Maciej Obremski. Non-malleable reductions and applications. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 459–468, 2015.

ADL14. Divesh Aggarwal, Yevgeniy Dodis, and Shachar Lovett. Non-malleable codes from additive combinatorics. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 774–783, 2014.

AGM+15. Shashank Agrawal, Divya Gupta, Hemanta K. Maji, Omkant Pandey, and Manoj Prabhakaran. A rate-optimizing compiler for non-malleable codes against bit-wise tampering and permutations. In *Theory of Cryptography - 12th Theory of Cryptography Conference, TCC 2015, Warsaw, Poland, March 23-25, 2015, Proceedings, Part I*, pages 375–397, 2015.

BBCM95. Charles H. Bennett, Gilles Brassard, Claude Crépeau, and Ueli M. Maurer. Generalized privacy amplification. *IEEE Transactions on Information Theory*, 41(6):1915–1923, 1995.

BBR88. Charles Bennett, Gilles Brassard, and Jean-Marc Robert. Privacy amplification by public discussion. *SIAM Journal on Computing*, 17(2):210–229, 1988.

CG14. Mahdi Cheraghchi and Venkatesan Guruswami. Non-malleable coding against bit-wise and split-state tampering. In *Theory of Cryptography - 11th Theory of Cryptography Conference, TCC 2014, San Diego, CA, USA, February 24-26, 2014. Proceedings*, pages 440–464, 2014.

CGL16. Eshan Chattopadhyay, Vipul Goyal, and Xin Li. Non-malleable extractors and codes, with their many tampered extensions. *IACR Cryptology ePrint Archive*, 2016:340, 2016.

CKOR10. Nishanth Chandran, Bhavana Kanukurthi, Rafail Ostrovsky, and Leonid Reyzin. Privacy amplification with asymptotically optimal entropy loss. In Leonard J. Schulman, editor, *STOC*, pages 785–794. ACM, 2010.

CL16. Eshan Chattopadhyay and Xin Li. Explicit non-malleable extractors, multi-source extractors and almost optimal privacy amplification protocols. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:36, 2016.

Coh16. Gil Cohen. Making the most of advice: New correlation breakers and their applications. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:52, 2016.

CRS11. Gil Cohen, Ran Raz, and Gil Segev. Non-malleable extractors with short seeds and applications to privacy amplification. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:96, 2011.

CZ14. Eshan Chattopadhyay and David Zuckerman. Non-malleable codes against constant split-state tampering. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 306–315, 2014.

DKRS06. Yevgeniy Dodis, Jonathan Katz, Leonid Reyzin, and Adam Smith. Robust fuzzy extractors and authenticated key agreement from close secrets. In Cynthia Dwork, editor, *Advances in Cryptology—CRYPTO 2006*, volume 4117 of *LNCS*, pages 232–250. Springer-Verlag, 20–24 August 2006.

DLWZ11. Yevgeniy Dodis, Xin Li, Trevor D. Wooley, and David Zuckerman. Privacy amplification and non-malleable extractors via character sums. In Rafail Ostrovsky, editor, *FOCS*, pages 668–677. IEEE, 2011.

DORS08. Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM Journal on Computing*, 38(1):97–139, 2008. arXiv:cs/0602007.

DPW10. Stefan Dziembowski, Krzysztof Pietrzak, and Daniel Wichs. Non-malleable codes. In *Innovations in Computer Science - ICS 2010, Tsinghua University, Beijing, China, January 5-7, 2010. Proceedings*, pages 434–452, 2010.

DS02. Y. Dodis and J. Spencer. On the (non-)universality of the one-time pad. In *43rd Annual Symposium on Foundations of Computer Science*, pages 376–385. IEEE, 2002.

DW09. Yevgeniy Dodis and Daniel Wichs. Non-malleable extractors and symmetric key cryptography from weak secrets. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, pages 601–610, Bethesda, Maryland, 31 May–2 June 2009.

GUV07. Venkatesan Guruswami, Christopher Umans, and Salil P. Vadhan. Unbalanced expanders and randomness extractors from Parvaresh-Vardy codes. In *IEEE Conference on Computational Complexity*, pages 96–108, 2007.

JKS93. Thomas Johansson, Gregory Kabatianskii, and Ben J. M. Smeets. On the relation between a-codes and codes correcting independent errors. In *Advances in Cryptology - EUROCRYPT '93, Workshop on the Theory and Application of of Cryptographic Techniques, Lofthus, Norway, May 23-27, 1993, Proceedings*, pages 1–11, 1993.

KOS17. Bhavana Kanukurthi, Sai Lakshmi Bhavana Obbattu, and Sruthi Sekar. Four-state non-malleable codes with explicit constant rate. In *Theory of Cryptography - 15th International Conference, TCC 2017, Baltimore, MD, USA, November 12-15, 2017, Proceedings, Part II*, pages 344–375, 2017.

KOS18. Bhavana Kanukurthi, Sai Lakshmi Bhavana Obbattu, and Sruthi Sekar. Non-malleable randomness encoders and their applications. In *Eurocrypt 2018, April 28-May 3*, 2018.

KR09. Bhavana Kanukurthi and Leonid Reyzin. Key agreement from close secrets over unsecured channels. In *Advances in Cryptology - EUROCRYPT 2009, 28th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Cologne, Germany, April 26-30, 2009. Proceedings*, pages 206–223, 2009.

Li12a. Xin Li. Design extractors, non-malleable condensers and privacy amplification. In *STOC*, pages 837–854, 2012.

Li12b. Xin Li. Non-malleable extractors, two-source extractors and privacy amplification. Cryptology ePrint Archive, Report 2012/188, 2012. `http://eprint.iacr.org/`.

Li15. Xin Li. Non-malleable condensers for arbitrary min-entropy, and almost optimal protocols for privacy amplification. In *Theory of Cryptography - 12th Theory of Cryptography Conference, TCC 2015, Warsaw, Poland, March 23-25, 2015, Proceedings, Part I*, pages 502–531, 2015.

Li17. Xin Li. Improved non-malleable extractors, non-malleable codes and independent source extractors. In *Symposium on Theory of Computing, STOC 2017, Montreal, Canada, June 19-23, 2017*, 2017.

Mau93. Ueli Maurer. Protocols for secret key agreement by public discussion based on common information. In Douglas R. Stinson, editor, *Advances in Cryptology—CRYPTO '93*, volume 773 of *LNCS*, pages 461–470. Springer-Verlag, 22–26 August 1993.

MW97. Ueli Maurer and Stefan Wolf. Privacy amplification secure against active adversaries. In Burton S. Kaliski, Jr., editor, *Advances in Cryptology—CRYPTO '97*, volume 1294 of *LNCS*, pages 307–321. Springer-Verlag, 1997.

NZ96. Noam Nisan and David Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52(1):43–53, 1996.

RW03. Renato Renner and Stefan Wolf. Unconditional authenticity and privacy from an arbitrarily weak secret. In Dan Boneh, editor, *Advances in Cryptology—CRYPTO 2003*, volume 2729 of *LNCS*, pages 78–95. Springer-Verlag, 2003.

A Appendix A

Proposition 1. *Let $A_1, ..., A_n$ be mutually exclusive and exhaustive events. Then, for probability distributions X_1, X_2 over some set S, we have:*

$$\mathbf{SD}\left(X_1, X_2\right) \leq \sum_{i=1}^{n} \Pr[A_i].\mathbf{SD}\left(X_1|A_i, X_2|A_i\right)$$

where $X_j|A_i$ is the distribution of X_j conditioned on the event A_i.

Proof.

$$\begin{aligned}
2\mathbf{SD}\left(X_1, X_2\right) &= \sum_{s \in S} \Bigg| \Pr[X_1 = s] - \Pr[X_2 = s] \Bigg| \\
&= \sum_{s \in S} \Bigg| \sum_{i=1}^{n} \Bigg(\Pr[A_i] \Pr[X_1 = s|A_i] - \Pr[A_i] \Pr[X_2 = s|A_i] \Bigg) \Bigg| \\
&\leq \sum_{s \in S} \sum_{i=1}^{n} \Pr[A_i] \Bigg| \Pr[X_1 = s|A_i] - \Pr[X_2 = s|A_i] \Bigg| \\
&= \sum_{i=1}^{n} \Pr[A_i] \sum_{s \in S} \Bigg| \Pr[X_1 = s|A_i] - \Pr[X_2 = s|A_i] \Bigg| \\
&= 2 \sum_{i=1}^{n} \Pr[A_i] \mathbf{SD}\left(X_1|A_i, X_2|A_i\right)
\end{aligned}$$

□

Proposition 2. *Let A, B be random variables over $\mathcal{A}$. Let C be any distribution over $\mathcal{C}$.*

$$\textit{If } \forall c \in \mathcal{C},\ \mathbf{SD}_{C=c}(A; B) \leq \epsilon \text{ , } \textit{then } \mathbf{SD}\left((A, C); (B, C)\right) \leq \epsilon$$

Proof.

$$\begin{aligned}
2\mathbf{SD}\left((A, C); (B, C)\right) &= \sum_{a,c} |\Pr[A = a, C = c] - \Pr[B = a, C = c]| \\
&= \sum_{c} \Pr[C = c] \sum_{a} |\Pr[A = a|C = c] - \Pr[B = a|C = c]| \\
&\leq \sum_{c} \Pr[C = c].\epsilon \\
&= \epsilon
\end{aligned}$$

□

Proposition 3. *Let $A_1, A_2, ..., A_n$ be mutually exclusive and exhaustive events. Let B be any (possibly correlated to A_i's) event with non-zero probability. Then*

$$\sum_{i=1 \ to \ n} \Pr[A_i|B] = 1$$

Proof.

$$\begin{aligned}\sum_{i=1 \ to \ n} \Pr[A_i|B] &= \sum_{i=1 \ to \ n} (\frac{\Pr[A_i \wedge B]}{\Pr[B]}) \\ &= \frac{\sum_{i=1 \ to \ n} \Pr[A_i \wedge B]}{\Pr[B]} \\ &= \frac{\Pr[B]}{\Pr[B]} \\ &= 1\end{aligned}$$

The third equation follows because A_i's are mutually exclusive and exhaustive events. □

Proposition 4. *Let A, B be random variables over $\mathcal{A}, \mathcal{B}$ respectively. Let F be some event with non-zero probability. Let C be the random variable $B|F$. Suppose A is independent of event F, then*

$$\forall a \in \mathcal{A}, b \in \mathcal{B}, \ \Pr[A = a, B = b|F] = \Pr[A = a, C = b] \ \textit{and}$$

$$\mathbf{SD}((A, B); (A, C)) \leq 1 - \Pr[F]$$

Proof. Define random variable D as $A|F$. Then

$$\begin{aligned}\Pr[A = a, B = b|F] &= \Pr[D = a, C = b] \\ &= \Pr[A = a, C = b]\end{aligned}$$

The above equation follows because A is independent of F and therefore, $D \equiv A$. Let $\tilde{F}$ be the complement event of F.

$$\begin{aligned}2\mathbf{SD}((A, B); (A, C)) &= \sum_{a,b} |\Pr[A = a, B = b] - \Pr[A = a, C = b]| \\ &= \sum_{a,b} |\Pr[F]\Pr[A = a, B = b|F] + \Pr[\tilde{F}]\Pr[A = a, B = b|\tilde{F}] - \Pr[A = a, C = b]| \\ &= \sum_{a,b} |\Pr[F]\Pr[A = a, C = b] + \Pr[\tilde{F}]\Pr[A = a, B = b|\tilde{F}] - \Pr[A = a, C = b]| \\ &leq \sum_{a,b} |\Pr[F]\Pr[A = a, C = b] - \Pr[A = a, C = b]| + \Pr[\tilde{F}]\sum_{a,b}\Pr[A = a, B = b|\tilde{F}] \\ &= (1 - \Pr[F])\sum_{a,b}(\Pr[A = a, C = b]) + \Pr[\tilde{F}]\Pr[A \in \mathcal{A}, B \in \mathcal{B}|\tilde{F}] \\ &= (1 - \Pr[F]) + \Pr[\tilde{F}].1 \\ &= 2(1 - \Pr[F])\end{aligned}$$

□

Proposition 5. *Let* A, B, C *be random variables and* F *be some event with non-zero probability. Suppose* A, C *are independent of* B*,* F *and* $A \approx_\epsilon C$*, then* $(A, B)|F \approx_\epsilon (C, B)|F$*.*

Proof. Let A', B', C' denote the random variables $(A|F), (B|F), (C|F)$. A, C are independent of B and F. Therefore, A', C' are independent of B'. For the sake of completeness we just show A' is independent of B'.

$$\begin{aligned}
\Pr[A' = a, B' = b] &= \Pr[A = a, B = b|F] \\
&= \Pr[A = a, B = b, F] / \Pr[F] \\
&= (\Pr[A = a] \Pr[B = b, F]) / \Pr[F] \\
&= \Pr[A = a] \Pr[B = b|F] \\
&= \Pr[A = a|F] \Pr[B = b|F] = \Pr[A' = a] \Pr[B' = b]
\end{aligned}$$

$$\begin{aligned}
2\mathbf{SD}\left((A, B)|F; (C, B)|F\right) &= \sum_{a,b} \Big| \Pr[A = a, B = b|F] - \Pr[C = a, B = b|F] \Big| \\
&= \sum_{a,b} \Big| \Pr[A' = a, B' = b] - \Pr[C' = a, B' = b] \Big| \\
&= \sum_{b} \Pr[B' = b] \sum_{a} \Big| \Pr[A' = a] - \Pr[C' = a] \Big| \\
&= \sum_{b} \Pr[B' = b] \sum_{a} \Big| \Pr[A = a] - \Pr[C = a] \Big| \\
&= \sum_{b} \Pr[B' = b] 2\epsilon = 2\epsilon
\end{aligned}$$

The above equations follow because A, C are independent of F and therefore, $A' \equiv A$ and $C' \equiv C$. □