A Message Franking Channel

Loïs Huguenin-Dumittan and Iraklis Leontiadis

LASEC, EPFL, Switzerland
Inpher, Switzerland and USA
{lois.huguenin-dumittan@epfl.ch, iraklis@inpher.io}

Abstract. We pursue to formalize and instantiate a secure bidirectional channel with message franking properties. Under this model a sender may send an abusive message to the receiver and the latter wish to open it in a verifiable way to a third party. Potential malicious behavior of a sender requires message franking protocols resistant to sending messages that cannot be opened later by the receiver. An adversary impersonated by the receiver may also try to open messages that have not been sent by the sender. Wrapping a message franking protocol in a secure channel requires a more delicate treatment in order to avoid drops or replay of messages and out-of-order delivery. To the best of our knowledge we are the first to model the security of a message franking channel, which apart from integrity, confidentiality, resistance to drops, relays and out-of-order delivery is *sender* and *receiver binding*: a sender cannot send a message which cannot be opened in a verifiable way later by the receiver, and the receiver cannot claim a message that had not been truly sent by the receiver. Finally, we instantiate a bidirectional message franking channel from symmetric primitives and analyze its security.

Keywords: message franking channel, secure communication, channel security, abusive verifiable reports

1 Introduction

The most popular messaging services such as Facebook Messenger, Whatsapp, Telegram or Signal offer end-to-end encryption, preventing anyone apart from the recipients from reading the messages. While preserving privacy, such schemes increase the difficulty of filtering spam or reporting abusive messages. Indeed, without the capacity to read the plaintexts, the router (e.g. Facebook) cannot check for abusive content, malware, malicious links, etc. The problem of abuse reporting was recently tackled by Facebook, which introduced the concept of *message franking* [Fac16]. With this proposed protocol, a user can report abusive messages to Facebook and can prove that an abusive message was sent by another user. More recently, Grubbs et al. initiated the formal analysis of such schemes [GLR17]. In particular, they introduced a new cryptographic primitive called committing authenticated encryption with associated data (committing AEAD) along with new security definitions. In the same paper, they analyze the security

of Facebook's scheme and present a more efficient construction. In a follow up work Dodis, Grubbs, Ristenpart and Woodage [DGRW18] revealed a flaw in Facebook message franking protocol for attachment delivery. The authors showed compromisation of sender binding, letting a malicious sender to send messages that cannot be reported. To circumvent that flaw they suggest a new design for message franking based on hashing encryption which provides the requiring commitment properties on the ciphertext with only one pass.

The security of communication protocols though depends not only on the underlying cryptographic primitives but also on the behavior of the protocol itself. For instance, a protocol based on secure primitives accepting out-of-order messages or an adversary being able to drop or replay messages renders communication between two end points vulnerable to such malicious behaviors. More generally, traditional cryptographic primitives cannot model real-world attacks beyond basic confidentiality and integrity. In particular, the integrity of the communication channel (e.g. security against out-of-order messages, message drops or replay attacks) is not captured by traditional security definitions. These reasons led to the study of communication protocols as stateful encryption schemes, so called: cryptographic channels [BKN02,KPB03,Mar17,PR18].

Cryptographic channel. Consider a messaging protocol where several participants share a key with each other and want to send and receive end-to-end encrypted messages. Once all the keys are fixed, a channel with confidentality, integrity–which includes resistance to replay attacks, out-of-order delivery and drops–must be established between each pair of participants. The cryptographic primitive that models the channels and the interaction between the participants is called a *cryptographic channel*. A channel where the only actions available to the clients are *send* and *receive* (i.e. as in a traditionnal message exchange protocol) will be referred to as a *standard cryptographic channel*.

A standard cryptographic channel $\mathsf{Ch} = (\mathsf{init}, \mathsf{snd}, \mathsf{rcv})$ is a tuple of three efficient algorithms that allows the participants to send and receive encrypted messages. If there are two participants and only one can send and only one can receive, the channel is unidirectional while if both can send and receive, the channel is bidirectional.

One could imagine that a bidirectional channel made up of two secure unidirectional channels should be secure. However, as shown in [MP17] this does not hold. In the same paper, the authors show several security results on bidirectional channels constructed from unidirectional channels in a special construction called the *canonic composition*. The more interesting ones concerning confidentality, are the following:

$$\text{IND-1CPA} + \text{IND-1CPA} \iff \text{IND-2CPA} \tag{1}$$

$$\text{IND-1CCA} + \text{IND-1CCA} \Longleftarrow \text{IND-2CCA} \tag{2}$$

$$\text{IND-1CCA} + \text{IND-1CCA} \not\Longrightarrow \text{IND-2CCA} \tag{3}$$

where IND-1CPA, IND-2CPA are the IND-CPA security for unidirectional and bidirectional channels, respectively, and IND-1CCA, IND-2CCA are the IND-CCA security for unidirectional and bidirectional channels, respectively

[MP17]. One can see that if a unidirectional channel is CPA secure, a CPA secure bidirectional channel can be constructed. On the other hand, two CCA secure unidirectional channels are not sufficient to create a CCA secure bidirectional channel. Intuitively, one can understand this result by considering the following example. We consider a bidirectional channel made up of two independent confidential unidirectional channels that do not guarantee integrity. Let the protocol be such that if an adversary sends a special ciphertext c' to Alice, she sends her password to Bob without a handshake. Note that this contradicts integrity but not condidentiality. Then, since Bob does not expect this message, he outputs the message (i.e. the password) in clear for everyone. Obviously, this does not contradict the confidentiality of the $B \to A$ channel. Now, we assume that Alice sends her password to Bob without a handshake if and only if she receives c'. Then, the $A \to B$ channel can still be confidential since unidirectional confidentiality (IND-1CCA) does not model the fact that Alice can receive messages in the $A \to B$ channel, in particular c'. Thus, these results show the importance of considering protocols in bidirectional cryptographic channels.

Recent results analyze security of secure communication channel over TLS [BHMS15,BH17,GM17], but without sender and receiver binding guarantees. In this work we aim to close the gap in the existing literature with the definitions, design and analysis of a secure communication channel with message franking properties: sender and receiver binding.

Our contributions. The contributions of this work are summarized as follows:

1. We first define a message franking channel (MFC) that models a messaging protocol where users can report abusive messages.
2. Then, we present unidirectional and bidirectional security definitions for our construction. The most challenging are the uni/bi-directional sender and receiver binding notions, which were introduced by Grubbs et al. [GLR17]. Specifically, binding definitions guarantee that a delivered message can be reported and a forged message cannot be reported.
3. We prove that a special construction called hereafter the *canonical composition*, made from two binding unidirectional MFC, is sufficient to build a secure binding bidirectional MFC.
4. Finally we present an instantiation of a message franking channel made from a secure committing AEAD scheme and a message authentication code.

Outline. In section 2 we introduce some notation for the manuscript. In section 3 we recap the reader the message franking protocol definitions and Facebook message franking protocol. We continue in section 4 with the syntactical definition of a message franking channel and in section 5 with the security properties thereof. A concrete instantiation is presented in section 6. Finally, we conclude our work in section 7.

2 Notation

A participant is referred to interchangeably as a client, a user or a party. We write $A \parallel B$ to denote the concatenation of A with B and $|A|$ to denote the length of A. We write $\Pr[G \Rightarrow x]$ to denote the probability a game G outputs x. If $\mathcal{X}$ is a set, then $X \leftarrow_{\$} \mathcal{X}$ means that X is uniformly sampled from $\mathcal{X}$. If G is a randomized algorithm, we write $x \leftarrow_{\$} G$ to denote the fact that x takes the value output by G. If G is deterministic, we write $x \leftarrow G$.
In the different games, we denote the initialization of an array A by $A \leftarrow []$. At each position i, an array can be assigned a single value x or a tuple of values $(x_1, \ldots, x_n)$. We denote these events by $A[i] \leftarrow x$ and $A[i] \leftarrow (x_1, \ldots, x_n)$, respectively. We write **abort** for "Stop the game and return 0 as a failure". If G is a game that returns a value n, we denote by $\Pr[G \Rightarrow n]$ the probability that G returns n.
Finally, when this is clear from the context, we denote by $*$ any value that could fit. For example, if $T \in \mathcal{X} \times \mathcal{Y}$ is a tuple of two values and $X \in \mathcal{X}$, we write $T == (X, *)$ to denote the event that there exists some $Y \in \mathcal{Y}$ such that T is equal to (X, Y).

3 Commiting AEAD

Grubbs et al. [GLR17] formalized the concept of message franking into the definition of *committing AEAD*. Roughly, the idea is to define a cryptographic primitive that creates a ciphertext and a commitment on the plaintext. Then, one can decrypt to retrieve the plaintext and an opening key, which is used to verify the commitement on the plaintext. We present here the definition of committing AEAD where all randomness is defined by a public nonce. Formally, a nonce-based committing AEAD is defined as follows:

Definition 1 (Nonce-based Committing AEAD). *A nonce-based committing AEAD scheme* $\mathsf{nCE} = (\mathsf{init}, \mathsf{enc}, \mathsf{dec}, \mathsf{vrf})$ *is a set of four algorithms. Associated to this scheme is a key space* $\mathcal{K}$*, a header space* $\mathcal{H}$*, a message space* $\mathcal{M}$*, a ciphertext space* $\mathcal{C}$*, an opening space* $\mathcal{K}_f$*, a franking tag space* $\mathcal{T}$ *and a nonce space* $\mathcal{N}$*. An error symbol* $\perp$ *is also required. The four algorithms are as follows:*

- $K \leftarrow_{\$} \mathsf{init}$*: The initialization algorithm* init *outputs a random key* $K \in \mathcal{K}$*.*
- $(C_1, C_2) \leftarrow \mathsf{enc}_K(N, H, M)$*: The encryption algorithm* enc *takes a key* $K \in \mathcal{K}$*, a nonce* $N \in \mathcal{N}$*, a header* $H \in \mathcal{H}$ *and a message* $M \in \mathcal{M}$*, and it outputs a ciphertext* $C_1 \in \mathcal{C}$ *and a franking tag* $C_2 \in \mathcal{T}$*.*
- $(M, K_f) \leftarrow \mathsf{dec}_K(N, H, C_1, C_2)$*: The decryption algorithm* dec *takes a key* $K \in \mathcal{K}$*, a nonce* $N \in \mathcal{N}$*, a header* $H \in \mathcal{H}$*, a ciphertext* $C_1 \in \mathcal{C}$ *and a franking tag* $C_2 \in \mathcal{T}$*, and it outputs a message* $M \in \mathcal{M}$ *and an opening* $K_f \in \mathcal{K}_f$ *or an error symbol* $\perp$*.*
- $b \leftarrow \mathsf{vrf}(H, M, K_f, C_2)$*: The verification algorithm* vrf *takes a header* $H \in \mathcal{H}$*, a message* $M \in \mathcal{M}$*, an opening value* $K_f \in \mathcal{K}_f$ *and a franking tag* $C_2 \in \mathcal{T}$*, and it outputs a verification bit* b*, regarding the correctness of the reporting procedure.*

The first procedure init is randomized while the others are deterministic, since the randomness is defined by the nonce value N.

Correctness. For the correctness of the scheme, we require that for any $K \in \mathcal{K}$, $N \in \mathcal{N}$, $H \in \mathcal{H}$ and $M \in \mathcal{M}$ and $(C_1, C_2) = \mathsf{enc}_K(N, H, M)$

$$\Pr[\mathsf{dec}_K(N, H, C_1) = (M, K_f)] = 1$$

for some $K_f \in \mathcal{K}_f$.
Also, if we let $(C_1, C_2) = \mathsf{enc}_K(N, H, M)$ and $(M, K_f) = \mathsf{dec}_K(N, H, C_1, C_2)$ then we require that

$$\Pr[\mathsf{vrf}(H, M, K_f, C_2) = 1] = 1$$

for any $K \in \mathcal{K}$, $N \in \mathcal{N}$, $H \in \mathcal{H}$ and $M \in \mathcal{M}$.
Finally, we require the length of the ciphertexts (C_1, C_2) to be deterministic given H and M. In other words, there is a deterministic function $len(H, M)$ s.t. $(|C_1|, |C_2|) = len(H, M)$.

Due to space constraints security definitions for AEAD are deferred in Appendix A.

4 Cryptographic channel for message franking (MFC)

A cryptographic channel is a set of algorithms that allows several participants to exchange messages (i.e. send and receive) with confidentiality, message integrity, resistance to replay attacks, out-of-order delivery and message drops. For the reasons exposed above and the fact that a bidirectional channel is more generic than a unidirectional one, we are going to focus on a bidirectional channel where two participants (Alice and Bob) exchange messages. However, in the message franking case, the participants do not only exchange messages but they also report them as abusive to a third entity, which we refer to as *router*. Therefore, we need to define a new model for a cryptographic channel, which we call a message franking channel (MFC). Informally, we raise the nonce-based committing AEAD concept to the channel level, in the context of message franking.

We define a message franking channel (MFC), using the syntax used by Marson et al. [MP17]:

Definition 2 (Message Franking Channel). *A message franking channel* $\mathsf{Ch} = (\mathsf{init}, \mathsf{snd}, \mathsf{tag}, \mathsf{rcv}, \mathsf{rprt})$ *is a five-tuple of algorithms. Associated to this channel is a key space* $\mathcal{K}$*, a nonce space* $\mathcal{N}$*, a header space* $\mathcal{H}$*, a message space* $\mathcal{M}$*, a ciphertext space* $\mathcal{C}$*, an opening space* $\mathcal{K}_f$*, a franking tag space* $\mathcal{T}$*, a router tag space* $\mathcal{T}_R$ *and a state space* $\mathcal{S}$*. The participants space is* $\mathcal{P} = \{A, B\}$ *(for Alice and Bob). We also require a special rejection symbol* $\perp \notin (\mathcal{K}_f \times \mathcal{M}) \cup \mathcal{S}$*. The five procedures are defined as follows:*

- $(st_A, st_B, st_R) \leftarrow_{\$} \mathsf{init}$*: The initialization algorithm* init *samples a key* $K \in \mathcal{K}$ *and a key* $K_R \in \mathcal{K}$*, and it outputs initial states* $st_A, st_B, st_R \in \mathcal{S}$*.* K *is the shared key resulting from a secure and authenticated key exchange protocol between both clients and* K_R *is the secret key of the router.*

- $(st'_u, H, C_1, C_2) \leftarrow \mathsf{snd}(st_u, N, M)$: *The sending algorithm* snd *takes the sender's state* $st_u \in \mathcal{S}$*, a nonce* $N \in \mathcal{N}$ *and a message* $M \in \mathcal{M}$*, and it outputs an updated state* $st'_u \in \mathcal{S}$*, a header* $H \in \mathcal{H}$ *and a pair of ciphertext and franking tag* $(C_1, C_2) \in \mathcal{C} \times \mathcal{T}$*.*
- $(st'_R, T_R) \leftarrow \mathsf{tag}(st_R, id_s, H, C_2)$: *The router tagging algorithm* tag *takes the router's state* $st_R \in \mathcal{S}$*, the sender's identity* $id_s \in \mathcal{P}$*, a header* $H \in \mathcal{H}$ *and a franking tag* $C_2 \in \mathcal{T}$*, and it outputs an updated state* $st'_R \in \mathcal{S}$ *and a router tag* $T_R \in \mathcal{T}_R$*.*
- $(st'_u, M, K_f) \leftarrow \mathsf{rcv}(st_u, N, H, C_1, C_2)$: *The receiving algorithm* rcv *takes the receiver's state* $st_u \in \mathcal{S}$*, a nonce* $N \in \mathcal{N}$*, a header* $H \in \mathcal{H}$ *, a ciphertext* $C_1 \in \mathcal{C}$ *and a franking tag* $C_2 \in \mathcal{T}$*, and it outputs an updated state* $st'_u \in \mathcal{S} \cup \{\perp\}$*, and an opening value pair* $(M, K_f) \in \mathcal{M} \times \mathcal{K}_f$ *or* $\perp$*. We require* $st'_u = \perp$ *if* $(M, K_f) = \perp$*.*
- $(st'_R, b) \leftarrow \mathsf{rprt}(st_R, id_r, H, M, K_f, C_2, T_R)$: *The router's verification algorithm* rprt *takes the router's state* $st_R \in \mathcal{S}$*, the reporter's identity* $id_r \in \mathcal{P}$*, a header* $H \in \mathcal{H}$*, a message* $M \in \mathcal{M}$*, an opening* $K_f \in \mathcal{K}_f$*, a franking tag* $C_2 \in \mathcal{T}$ *and a router tag* $T_R \in \mathcal{T}_R$*, and it outputs an updated router's state* $st'_R \in \mathcal{S} \cup \{\perp\}$ *and a verification bit* $b \in \{0, 1\}$*. We require* $st'_R = \perp$ *if* $b = 0$*.*

We assume the channel is stateful, i.e. the participants save their state between send/tag/receive/report calls. The rcv procedure can verify the commitment C_2 and outputs $\perp$ if the verification fails. Also, note that the rprt procedure does not depend on the nonce N nor on the ciphertext C_1. In particular, this means that the router must be able to verify the validity of the router tag T_R given only its state st_R, the header H and the franking tag C_2. The channel uses $\perp$ to indicate an error. Once a state is marked as bogus (i.e. $st = \perp$), it cannot be used anymore in the invokation of the functions snd, rcv and rprt, since $\perp \notin S$. This corresponds to the reasonable behaviour of an application refusing to process any more data once an error has been detected. Error management in channels is a full topic in itself (e.g. [BDPS13]) and can lead to vulnerabilities (e.g. padding oracle attack [Vau02]). Here, we assume that an adversary does not learn anything from an error apart from the failure of the corresponding procedure. The randomness is uniquely determined by public nonces N. Nonces can be used to perform randomized encryption but also to generate a random opening key K_f, as in Facebook's scheme. The role of the nonce is determined by the underlying schemes used by the different algorithms.

Remarks. The key K_R is the router's secret key that can be used to generate the router tags T_R. In a real-life message franking protocol, all messages go through the router, where they get tagged. Otherwise, clients could bypass the tagging procedure and messages would not be reportable. Therefore, one could imagine that a unique procedure for sending and tagging would be sufficient. While making the MFC definition simpler, this would render the instantiation of a real MFC difficult. Indeed, snd is meant to be executed by a client while tag is run by the router. Thus, such a simplification would be impractical. However, in the oracles of adversarial models used in the following sections, snd and tag

procedures will sometimes be considered as one operation, to model the fact that a message sent is always seen by the router.

It is important for the security of the scheme, in particular for the receiver-binding property, that the router knows the identity of the sender and the identity of the receiver. This is why these identities id_s and id_r are passed as arguments in the tag and rprt procedures, respectively. Obviously, the router should be able to verify these identities in order for the whole protocol to be secure, otherwise one could tag messages on behalf of another user. However, throughout this paper we assume that the parties and the router have established secure keys and that the router can authenticate the sender and the reporter in the corresponding procedures. It is a fair assumption since messaging protocols usually encrypt and authenticate the communications between a client and the server. For example, TextSecure (Signal's ancestor) used to encrypt client-to-router communications with TLS and it used to authenticate the client with the phone number concatenated with some secret key [FMB+14]. Whatsapp encrypts communications to its servers with the Noise Protocol Framework and it stores the client's Curve25519 public key, allowing the router to authenticate the user during the Diffie-Hellman key exchange protocol [Per18,Wha17].

4.1 Correctness of the channel.

As in a standard bidirectional channel [MP17], we require a MFC to have certain properties. In short, we want messages sent by a client to be decryptable by the other participant without errors, assuming that the ciphertexts are not modified in the channel and that the order of the messages is preserved. Also, a message sent by a honest participant in the channel and correctly deciphered at the receiving end should be reportable, assuming that all messages sent on the channel are not modified (i.e. no active adversary).

Such requirements can be represented as a game, where only passive external adversaries (i.e. adversaries that can only see and relay messages) are allowed and where the participants are honest. The adversary can schedule snd/tag, rcv and rprt procedures and it wins if a message sent can not be decrypted or reported (i.e. rprt fails). This game is represented in Figure 1. We assume $u \in \{A, B\}$, $N \in \mathcal{N}$, $H \in \mathcal{H}$, $M \in \mathcal{M}$, $C_1 \in \mathcal{C}$, $C_2 \in \mathcal{T}$, $K_f \in \mathcal{K}_f$ and $T_R \in \mathcal{T}_R$. The variables s_A, s_B, r_A, r_B keep track of the number of messages sent and received by each participant. The variables h_A and h_B keep track of the state of each participant (whether it has received modified/out-of-order messages). Note that an adversary can be external and try to modify the messages on the channel or it can be a participant who communicates with the other benign participant and/or the router. If a participant u receives out-of-order/modified data, (i.e. if the adversary actively attacks u), its variable h_u is set to false. If $h_u = true$ we say that u is clean. Observe that the adversary can win only if the participant u used in the oracle query is clean. $M_{A,B}[]$, $M_{B,A}[]$, $C_{A,B}[]$ and $C_{B,A}[]$ record the messages and ciphertexts exchanged. The adversary has access to three oracles:

- $\mathcal{O}^{snd}(u, N, M)$: The $\mathcal{O}^{snd}$ oracle takes the client identity u, a message and a nonce, and it calls the snd and tag procedures on behalf of the participant

u. Then, it records the resulting ciphertext and franking/router tags, if the client is still clean. This prevents the adversary from winning if the client u had previously received out-of-order/modified messages.

- $\mathcal{O}^{rcv}(u, N, H, C_1, C_2)$: With the oracle $\mathcal{O}^{rcv}$, an adversary can make a user u receive (i.e. decrypt). Now, if this participant is still clean and the ciphertext is delivered without modification (and in the right order) compared to the one sent (condition in line 4), then the message recovered should be the one sent. If this is not the case (condition in line 6), then the adversary wins. Otherwise, if the condition at line 4 is not respected, then the ciphertext/tags have been modified and the participant is flagged as not clean.
- $\mathcal{O}^{rprt}(u, n, H, M, K_f, C_2, T_R)$: The adversary can use also the $\mathcal{O}^{rprt}$ oracle to report a given message on behalf of a user u, by providing the message along with the tags. It also must specify the index n of the message (we assume the adversary records the number of messages sent). As before, if the user is clean and the header, the franking tag, the router tag and the message were not modified compared to the one sent (conditions at lines 2-3), then the participant/adversary should be able to report the message. If this is not the case, the adversary wins. Otherwise, if the condition in line 3 is not respected (i.e. the message, ciphertext or tags have been tampered with), the participant is flagged as not clean.

Now, for any adversary $\mathcal{A}$ playing CORR and any channel Ch, we denote the advantage of $\mathcal{A}$ as

$$\mathsf{Adv}^{\mathrm{corr}}_{\mathsf{Ch}}(\mathcal{A}) = \Pr[\mathrm{CORR}_{\mathsf{Ch}}(\mathcal{A}) \Rightarrow \mathsf{true}]$$

Definition 3 (Bidirectional Message Franking Channel Correctness). *We say that a message franking bidirectional channel* Ch *is correct if for any adversary* $\mathcal{A}$ *playing the CORR game*

$$\mathsf{Adv}^{\mathrm{corr}}_{\mathsf{Ch}}(\mathcal{A}) = 0$$

In Appendix B, we give the unidirectional correctness definition.

5 Security for message franking channel

5.1 Confidentiality

We elevate the MO-nRAND confidentiality game for nonce-based committing AEAD schemes (see Section A) to the bidirectional MFC case. Confidentiality means that no adversary is able to retrieve information about the messages exchanged by the participants. This notion concerns the exchange of messages and not the reporting phase. In addition, we note that once an abusive message is

$\mathrm{CORR}_{\mathsf{Ch}}(\mathcal{A})$

$win \leftarrow \mathsf{false}$
$(st_A, st_B, st_R) \leftarrow\$ \mathsf{init}$
$s_A \leftarrow 0; s_B \leftarrow 0$
$r_A \leftarrow 0; r_B \leftarrow 0$
$h_A \leftarrow \mathsf{true}; h_B \leftarrow \mathsf{true}$
$M_{A,B} \leftarrow []; M_{B,A} \leftarrow []$
$C_{A,B} \leftarrow []; C_{B,A} \leftarrow []$
$\mathcal{A}^{\mathcal{O}^{snd}, \mathcal{O}^{rcv}, \mathcal{O}^{rprt}}$
return win

Oracle $\mathcal{O}^{rprt}(u, n, H, M, K_f, C_2, T_R)$

$(st_R, b) \leftarrow \mathsf{rprt}(st_R, u, H, M, K_f, C_2, T_R)$
if h_u :
  if $C_{v,u}[n] = (*, H, *, C_2, T_R)$ and $M_{v,u}[n] = M$:
    if $b = 0$:
      $win \leftarrow \mathsf{true}$
  else :
    $h_u \leftarrow \mathsf{false}$
return b

Oracle $\mathcal{O}^{snd}(u, N, M)$

$(st_u, H, C_1, C_2) \leftarrow \mathsf{snd}(st_u, N, M)$
$(st_R, T_R) \leftarrow \mathsf{tag}(st_R, u, H, C_2)$
$v \leftarrow \{A, B\} \setminus \{u\}$
if h_u :
  $M_{u,v}[s_u] \leftarrow M$
  $C_{u,v}[s_u] \leftarrow (N, H, C_1, C_2, T_R)$
  $s_u \leftarrow s_u + 1$
return (C_1, C_2, T_R)

Oracle $\mathcal{O}^{rcv}(u, N, H, C_1, C_2)$

$(st_u, M, K_f) \leftarrow \mathsf{rcv}(st_u, N, H, C_1, C_2)$
$v \leftarrow \{A, B\} \setminus \{u\}$
if h_u :
  if $r_u < s_u$ and $C_{v,u}[r_u] = (N, H, C_1, C_2, *)$:
    if $(M, K_f) = \perp$ or $M_{v,u}[r_u] \neq M$:
      $win \leftarrow \mathsf{true}$
    $r_u \leftarrow r_u + 1$
  else :
    $h_u \leftarrow \mathsf{false}$
return (M, K_f)

Fig. 1. Correctness game for a bidirectional message franking channel.

reported, it becomes public, in the sense that we do not specify how the message is sent to the router for reporting (e.g. it could be sent in clear).

While in a standard channel the adversary is external to the participants, in the message franking case the adversary can also be the router. Therefore, in the following games we give the adversary access to the router's state st_R. This means the adversary does not need a tag and report oracle, as it is able to run these procedures on its own.

The MO-2nREAL and MO-2nRAND games of Figure 2 are adapted from the MO-nREAL and MO-nRAND games of Section A. In both games, we assume the adversary is nonce-respecting (i.e. it cannot query $\mathcal{O}^{snd}$ or $\mathcal{O}^{chal}$ twice with the same nonce N.

The adversary wins if it can differentiate with non-negligible probability between the encryption of a message M and the encryption of a random bitstring. The h_u variables, as in the correctness games, let the adversary decrypt ciphertexts as long as it remains passive: the adversary only relays messages. In particular, this prevents the adversary from winning by decrypting the ciphertexts obtained from the $\mathcal{O}^{chal}$ oracle.

$\text{MO-2nREAL}_{\mathsf{Ch}}(\mathcal{A})$

$(st_A, st_B, st_R) \leftarrow\$ \mathsf{init}$
$s_A \leftarrow 0; s_B \leftarrow 0; r_A \leftarrow 0; r_B \leftarrow 0$
$h_A \leftarrow \mathsf{true}; h_B \leftarrow \mathsf{true}$
$C_{A,B} \leftarrow []; C_{B,A} \leftarrow []$
$b' \leftarrow \mathcal{A}^{\mathcal{O}^{snd}, \mathcal{O}^{rcv}, \mathcal{O}^{chal}, st_R}$
return b'

Oracle $\mathcal{O}^{snd}(u, N, M)$

$(st_u, H, C_1, C_2) \leftarrow \mathsf{snd}(st_u, N, M)$
$v \leftarrow \{A, B\} \setminus \{u\}$
if h_u :
  $C_{u,v}[s_u] \leftarrow (N, H, C_1, C_2)$
  $s_u \leftarrow s_u + 1$
return (H, C_1, C_2)

Oracle $\mathcal{O}^{rcv}(u, N, H, C_1, C_2)$

$(st_u, M, K_f) \leftarrow \mathsf{rcv}(st_u, N, H, C_1, C_2)$
$v \leftarrow \{A, B\} \setminus \{u\}$
if $r_u < s_v$ and $C_{v,u}[r_u] = (N, H, C_1, C_2)$:
  $r_u \leftarrow r_u + 1$
else :
  $h_u \leftarrow \mathsf{false}$
if h_u: **return** (M, K_f)
else : **return** $\perp$

Oracle $\mathcal{O}^{chal}(u, N, M)$

$(st_u, H, C_1, C_2) \leftarrow \mathsf{snd}(st_u, N, M)$
return (C_1, C_2)

$\text{MO-2nRAND}_{\mathsf{Ch}}(\mathcal{A})$

$(st_A, st_B, st_R) \leftarrow\$ \mathsf{init}$
$s_A \leftarrow 0; s_B \leftarrow 0; r_A \leftarrow 0; r_B \leftarrow 0$
$h_A \leftarrow \mathsf{true}; h_B \leftarrow \mathsf{true}$
$C_{A,B} \leftarrow []; C_{B,A} \leftarrow []$
$b' \leftarrow \mathcal{A}^{\mathcal{O}^{snd}, \mathcal{O}^{rcv}, \mathcal{O}^{chal}, st_R}$
return b'

Oracle $\mathcal{O}^{snd}(u, N, M)$

$(st_u, H, C_1, C_2) \leftarrow \mathsf{snd}(st_u, N, M)$
$v \leftarrow \{A, B\} \setminus \{u\}$
if h_u :
  $C_{u,v}[s_u] \leftarrow (N, H, C_1, C_2)$
  $s_u \leftarrow s_u + 1$
return (H, C_1, C_2)

Oracle $\mathcal{O}^{rcv}(u, N, H, C_1, C_2)$

$(st_u, M, K_f) \leftarrow \mathsf{rcv}(st_u, N, H, C_1, C_2)$
$v \leftarrow \{A, B\} \setminus \{u\}$
if $r_u < s_v$ and $C_{v,u}[r_u] = (N, H, C_1, C_2)$:
  $r_u \leftarrow r_u + 1$
else :
  $h_u \leftarrow \mathsf{false}$
if h_u: **return** (M, K_f)
else : **return** $\perp$

Oracle $\mathcal{O}^{chal}(u, N, M)$

$M_r \leftarrow\$ \{0,1\}^{|M|}$
$(st_u, H, C_1, C_2) \leftarrow \mathsf{snd}(st_u, N, M_r)$
return (C_1, C_2)

Fig. 2. Confidentality games for nonce-based bidirectional message franking channels.

For a bidirectional MFC channel Ch, we define the nonce-based multiple opening real-or-random (MO-2nRoR) advantage of any algorithm $\mathcal{A}$ as

$$\mathsf{Adv}^{\text{mo-2nror}}_{\mathsf{Ch}}(\mathcal{A}) = |\Pr[\text{MO-2nREAL}_{\mathsf{Ch}}(\mathcal{A}) \Rightarrow 1] - \Pr[\text{MO-2nRAND}_{\mathsf{Ch}}(\mathcal{A}) \Rightarrow 1]|$$

5.2 Integrity

We adapt the MO-nCTXT integrity notions of committing AEAD to the bidirectional MFC. Ciphertext integrity means that a receiver can only receive and decrypt in-order and legitimate ciphertexts, which have been sent by another

$\text{MO-2nCTXT}_{\mathsf{Ch}}(\mathcal{A})$

$(st_A, st_B, st_R) \leftarrow\$ \,\mathsf{init}$
$s_A \leftarrow 0; s_B \leftarrow 0; r_A \leftarrow 0; r_B \leftarrow 0$
$C_{A,B} \leftarrow []; C_{B,A} \leftarrow []$
$win \leftarrow \mathsf{false}$
$\mathcal{A}^{\mathcal{O}^{snd}, \mathcal{O}^{rcv}, st_R}$
return win

Oracle $\mathcal{O}^{snd}(u, N, M)$

$v \leftarrow \{A, B\} \setminus \{u\}$
$(st_u, H, C_1, C_2) \leftarrow \mathsf{snd}(st_u, N, M)$
$C_{u,v}[s_A] \leftarrow (N, H, C_1, C_2)$
$s_u \leftarrow s_u + 1$
return (H, C_1, C_2)

Oracle $\mathcal{O}^{rcv}(u, N, H, C_1, C_2)$

$(st_u, M, K_f) \leftarrow \mathsf{rcv}(st_u, N, H, C_1, C_2)$
$v \leftarrow \{A, B\} \setminus \{u\}$
if $(M, K_f) = \perp$: **return** $\perp$
if $r_u < s_v$ and $(N, H, C_1, C_2) = C_{v,u}[r_u]$:
  $r_u \leftarrow r_u + 1$
  return (M, K_f)
else : $win \leftarrow \mathsf{true}$

Fig. 3. Integrity game for nonce-based committing AEAD

participant. As in the MO-2nROR confidentiality notion, the adversary can be the router and thus we give access to the state st_R.

Let MO-2nCTXT be the game in Figure 3. By observing the $\mathcal{O}^{rcv}$ oracle, one can see that the adversary wins if the rcv procedure is successful but the ciphertexts are not legitimate or they are received out-of-order. For a bidirectional MFC Ch, we define the nonce-based multiple opening integrity advantage of any algorithm $\mathcal{A}$ playing the MO-2nCTXT game as

$$\mathsf{Adv}^{\text{mo-2nctxt}}_{\mathsf{Ch}}(\mathcal{A}) = \Pr[\text{MO-2nCTXT}_{\mathsf{Ch}}(\mathcal{A}) \Rightarrow \mathsf{true}]$$

5.3 Binding security notions

In order to guarantee verifiable reporting of abusive messages, a MFC should adhere to sender and receiver binding notions as with message franking protocols [Fac16,GLR17]. One difference between the committing AEAD binding definitions and the ones defined here, is that in a channel the participants are stateful and their behavior may evolve over time, whereas such concepts do not exist in the security definitions of cryptographic primitives like committing AEAD. The threat in such security definitions is a malicious participant and not an external adversary. Therefore, in all the binding games of this section, we assume that the adversary can modify the states of the participants (Alice and Bob). In particular, this allows the adversary to control sthe encryption key K.

Sender binding. Sender binding for MFC guarantees that any message received without error by a client can be successfully reported to the router. This notion is defined with the game s-2BIND presented in Figure 4.

The adversary has access to three oracles in addition to the states of Alice and Bob. Here is a description of the s-2BIND game:

s-2BIND$_{\mathsf{Ch}}(\mathcal{A})$

1 : $win \leftarrow \mathsf{false}$
2 : $(st_A, st_B, st_R) \leftarrow\$ \mathsf{init}$
3 : $s_A \leftarrow 0; s_B \leftarrow 0$
4 : $r_A \leftarrow 0; r_B \leftarrow 0$
5 : $S_{A,B} \leftarrow []; S_{B,A} \leftarrow []$
6 : $R_{A,B} \leftarrow []; R_{B,A} \leftarrow []$
7 : $\mathcal{A}^{\mathcal{O}^{tag}, \mathcal{O}^{rcv}, \mathcal{O}^{rprt}, st_A, st_B}$
8 : **return** win

Oracle $\mathcal{O}^{rprt}(u, H, M, K_f, C_2, T_R)$

1 : $id_r \leftarrow u$
2 : $v \leftarrow \{A, B\} \setminus \{u\}$
3 : $(st_R, b) \leftarrow \mathsf{rprt}(st_R, id_r, H, M, K_f, C_2, T_R)$
4 : **if** (H, M, K_f, C_2, T_R) in $R_{v,u}$ and $b = 0$:
5 : $\quad win \leftarrow \mathsf{true}$
6 : **return** b

Oracle $\mathcal{O}^{rcv}(u, N, H, C_1, C_2)$

1 : $v = \{A, B\} \setminus \{u\}$
2 : **if** (H, C_2, T_R) in $S_{v,u}$ for some T_R :
3 : $\quad (st_u, M, K_f) \leftarrow rcv(st_u, N, H, C_1, C_2)$
4 : $\quad$ **if** $(M, K_f) \neq \perp$
5 : $\quad\quad R_{v,u}[r_u] \leftarrow (H, M, K_f, C_2, T_R)$
6 : $\quad\quad r_u \leftarrow r_u + 1$

Oracle $\mathcal{O}^{tag}(u, H, C_2)$

1 : $id_s \leftarrow u$
2 : $(st_R, T_R) \leftarrow \mathsf{tag}(st_R, id_s, H, C_2)$
3 : $v = \{A, B\} \setminus \{u\}$
4 : $S_{u,v}[s_u] = (H, C_2, T_R)$
5 : $s_u \leftarrow s_u + 1$
6 : **return** T_R

Fig. 4. Game for message franking channel sender-binding security definition.

- s-2BIND$_{\mathsf{Ch}}(\mathcal{A})$: The r_A, r_B, s_A, s_B variables keep track of the number of messages received and sent by each party. The arrays $S_{A,B}[], S_{B,A}[]$ store the franking tags C_2, the headers H and the associated router tags T_R sent from A to B and from B to A, respectively. The arrays $R_{A,B}[]$ and $R_{B,A}[]$ store the tags and messages corresponding to ciphertexts when the rcv procedure outputs a valid pair (M, K_f), after decryption.
- $\mathcal{O}^{tag}(u, H, C_2)$: This oracle is used to modify the router state st_R and to obtain a router tag for a tuple (H, C_2), with u as the sender. The header H, the franking tag C_2 and the router tag obtained T_R are stored in the corresponding array $S_{u,v}[]$ (line 4). This array is needed to ensure that a router tag T_R really corresponds to a header,franking tag pair (H, C_2). Then, the number of messages sent by the participant is incremented (line 5). Finally, the router tag is returned to the adversary (line 6).
- $\mathcal{O}^{rcv}(u, N, H, C_1, C_2)$: The $\mathcal{O}^{rcv}$ oracle first checks that H and C_2 corresponds to a router tag T_R (line 2) computed for a message sent by client v. If this is the case, the rcv procedure is called and it outputs a plaintext, opening key pair (M, K_f). If this pair is valid (i.e. not equal to $\perp$), it is stored in the appropriate array $R_{v,u}$ alongside with the header and the tags (C_2, T_R) (line 5). This means that these values correspond to a ciphertext that decrypts to a valid message. Finally, the number of valid messages received by u is incremented (line 6).
- $\mathcal{O}^{rprt}(u, H, M, K_f, C_2, T_R)$: The oracle $\mathcal{O}^{rprt}$ calls the rprt procedure on the the header, the message, the opening key and the tags, with u as the reporter. This allows an adversary to update the router state st_R. Then, if the input

r-2BIND$_{\mathsf{Ch}}(\mathcal{A})$	**Oracle** $\mathcal{O}^{snd}(u, N, H, M)$
1 : $win \leftarrow \mathsf{false}$	1 : $id_s \leftarrow u$
2 : $(st_A, st_B, st_R) \leftarrow\$ \mathsf{init}$	2 : $(st_u, _, C_1, C_2) \leftarrow \mathsf{snd}(st_u, N, M)$
3 : $s_A \leftarrow 0; s_B \leftarrow 0$	3 : $(st_R, T_R) \leftarrow \mathsf{tag}(st_R, id_s, H, C_2)$
4 : $M_{A,B} \leftarrow []; M_{B,A} \leftarrow []$	4 : $v = \{A, B\} \setminus \{u\}$
5 : $\mathcal{A}^{\mathcal{O}^{snd}, \mathcal{O}^{rprt}, st_A, st_B}$	5 : $M_{u,v}[s_u] \leftarrow M$
6 : **return** win	6 : $s_u \leftarrow s_u + 1$
	7 : **return** (C_1, C_2, T_R)

Oracle $\mathcal{O}^{rprt}(u, H, M, K_f, C_2, T_R)$

1 : $id_r \leftarrow u; v = \{A, B\} \setminus \{u\}$
2 : $(st_R, b) \leftarrow \mathsf{rprt}(st_R, id_r, H, M, K_f, C_2, T_R)$
3 : **if** M **not** in $M_{v,u}$ and $b = 1$:
4 : $win \leftarrow \mathsf{true}$
5 : **return** b

Fig. 5. Game for franking channel receiver-binding security definition.

values correspond to some ciphertext that outputs a valid message but the report procedure fails (line 4), the adversary wins.

The adversary can use the $\mathcal{O}^{rprt}$ oracle to report a message on behalf of a user u. If the message was actually sent and correctly decrypted by u then it is in $R_{v,u}$ with the corresponding header and tags. Therefore, if the message is in $R_{v,u}$ but the rprt procedure fails, the adversary wins. Indeed, we ask an adversary to win if the decryption is successful but the report procedure fails. We need to use a second pair of arrays $(S_{A,B}, S_{B,A})$ to store the router tag T_R corresponding to a pair (H, C_2). Otherwise, one could not store T_R in the array $R_{v,u}$ at line 5 of $\mathcal{O}^{rcv}$. Indeed, if T_R was not stored in $R_{v,u}$, the adversary could specify a random T_R when reporting with legit values, making the verification fail. Finally, we note that despite the fact the adversary has access to both user states, in a real-life threat model the adversary would be only one of the participants (e.g. Alice). However, since both participants share everything, it does not matter if the adversary has access to one or both states. For any adversary $\mathcal{A}$ playing s-2BIND and any channel Ch, the advantage of $\mathcal{A}$ is:

$$\mathsf{Adv}^{\mathrm{s-2bind}}_{\mathsf{Ch}}(\mathcal{A}) = \Pr[\text{s-2BIND}_{\mathsf{Ch}}(\mathcal{A}) \Rightarrow \mathsf{true}]$$

Receiver binding. We recall that receiver binding assures that a malicious participant cannot report an abusive message that was never sent. Receiver binding for message franking channels is defined with the game r-2BIND$_{\mathsf{Ch}}$ presented in Figure 5. In this game, the adversary represents two colluding participants that can schedule $\mathsf{snd}/\mathsf{tag}/\mathsf{rprt}$ operations for message exchanges and then one of the participants tries to report a message never sent to her/him. Since the adversary controls the users states, it would need only access to $\mathsf{tag}/\mathsf{rprt}$ oracles since it can run the other algorithms by itself. However, the game should store the sent messages (i.e. the ones the adversary requested a router tag for) in order

to compare them with the reported message in the $\mathcal{O}^{rprt}$ oracle. Therefore, we let the $\mathcal{O}^{snd}$ oracle run the snd procedure as well as the tag procedure. Here is a description of the game:

- r-2BIND$_{\mathsf{Ch}}(\mathcal{A})$: The s_A, s_B variables keep track of the number of messages sent by each party. They are used as indexes for the arrays $M_{A,B}[], M_{B,A}[]$ that store all messages sent by A to B and B to A, respectively.
- $\mathcal{O}^{snd}(u, N, H, M)$: In the $\mathcal{O}^{snd}$ oracle, the message sent M is recorded in the corresponding array (line 5) and the number of messages sent by the participant is incremented (line 6). The ciphertext, franking tag and router tag are returned to the adversary (line 7). In short, this oracle allows an adversary to obtain the router tag for any header/message tuple, while recording the message. We let the adversary pass the header H as an argument since we want the adversary to be able to get a router tag for any header.
- $\mathcal{O}^{rprt}(u, H, M, K_f, C_2, T_R)$: The $\mathcal{O}^{rprt}$ takes a header H, the message to be reported M, an opening key K_f, a franking tag C_2, a router tag T_R and the identity of the reporter u. At line 3, the condition checks that the submitted message was not sent to the reporting participant and that the message passes the rprt procedure. If this is the case, the adversary wins.

In order to respect this receiver binding notion, the router must be able to check the sender and receiver of a message. Indeed, if this is not the case, one can construct an adversary $\mathcal{A}^{\text{r-2BIND}}$ that always wins, as shown in Figure 6. Let Alice be the malicious participant. She picks a random message and a random nonce and sends the corresponding ciphertext to Bob using the sending oracle. Thus, the message is in the array $M_{A,B}$. Note that Alice can compute the opening key K_f since she controls st_A, N, H and M. Finally, she reports Bob as the sender of the message by calling the $\mathcal{O}^{rprt}$ oracle. Since the message is in $M_{A,B}$ but not in $M_{B,A}$, the first part of the condition in line 3 of $\mathcal{O}^{rprt}$ is fulfilled. Thus, since all values used to report are legitimate, the report procedure will succeed unless the router knows the message was actually sent by Alice to Bob.

One solution to this problem is to incorporate the receiver/sender identities in the router tag T_R. This is done in the Facebook protocol by putting the sender and receiver identities in some context data. The router tag T_R becomes $T_R = \mathsf{HMAC}_{K_R}(C_2 \parallel sender \parallel receiver)$ and the router can check whether the participant A reporting an abusive message from participant A is telling the truth or not and accept or reject accordingly. This shows the importance to correctly manage the identity of the participants.

For any adversary $\mathcal{A}$ playing r-2BIND and any channel Ch, we write the advantage of $\mathcal{A}$ as

$$\mathsf{Adv}^{\text{r}-2\text{bind}}_{\mathsf{Ch}}(\mathcal{A}) = \Pr[\text{r-2BIND}_{\mathsf{Ch}}(\mathcal{A}) \Rightarrow \mathsf{true}]$$

In other words, in addition to the control of participants' states, the adversary must be able to modify the state of the router, as in a real exchange of messages.

In Appendix B, we give the unidirectional binding definitions.

$\mathcal{A}^{\text{r-2BIND}}$

1 : $M \leftarrow\$ \mathcal{M}; N \leftarrow\$ \mathcal{N}$
2 : compute next K_f from (N, H, M, st_A)
3 : $(H, C_1, C_2, T_R) \leftarrow \mathcal{O}^{snd}(A, N, M)$
4 : $\mathcal{O}^{rprt}(A, H, M, K_f, C_2, T_R)$

Fig. 6. Attack on r-BIND game.

6 MFC Instantiation

Now that the security properties of a MFC have been defined, we wish to instantiate a practical MFC that fulfills these properties. Our construction is based on two cryptographic primitives, namely a committing AEAD and a message authentication code (MAC) scheme. We show how to combine these primitives with some message counters to obtain a practical bidirectional MFC.

6.1 Construction

Let $\mathsf{nCE} = (\mathsf{init}_{\mathsf{nCE}}, \mathsf{enc}, \mathsf{dec}, \mathsf{vrfy}_{\mathsf{nCE}})$ be a secure nonce-based committing AEAD and $\mathsf{MAC} = (\mathsf{init}_{\mathsf{mac}}, \mathsf{tag}_{\mathsf{mac}}, \mathsf{vrfy}_{\mathsf{mac}})$ be a secure MAC scheme. Our MFC construction is given in Figure 7. We refer to it as MF. The channel operates as follows:

- $(st_A, st_B, st_R) \leftarrow\$ \mathsf{init}$: The initialization procedure samples the keys and creates the states. Alice and Bob states are made of their identity, the secret key and the send and receive counters. The router's state is made of the router key.
- $(st, H, C_1, C_2) \leftarrow \mathsf{snd}(st, N, M)$: The sending procedure computes the header H as the identity concatenated with the number of sent messages. Then, the ciphertexts are computed, the sent counter is incremented and the values are returned.
- $(st_R, T_R) \leftarrow \mathsf{tag}(st_R, id_s, H, C_2)$: The router tag is computed as a MAC on the sender/receiver identities, the header and the franking tag, with the router key K_R.
- $(st, M, K_f) \leftarrow \mathsf{rcv}(st, N, H, C_1, C_2)$: The sender identity and the message sequence number are extracted from the header and the ciphertext is decrypted. If this fails or if the number of received messages is not equal to the message sequence number, an error is returned. Otherwise, the number of received messages is incremented and the new state, the plaintext and the opening key are returned.
- $(st_R, b) \leftarrow \mathsf{rprt}(st_R, id_r, H, M, K_f, C_2, T_R)$: The reporter's identity u and the alleged sender's identity v are extracted. Then, the router's tag T_R is verified with the identities v and u along with the header H and the franking tag C_2. Finally, the message is verified to be legit using the franking tag and the opening key K_f. If everything is successful, the procedure returns 1, as a success.

```
init
K ←$ 𝒦; K_R ←$ 𝒦
s_A ← 0; s_B ← 0
r_A ← 0; r_B ← 0
st_A ← A||K||s_A||r_A
st_B ← B||K||s_B||r_B
st_R ← K_R
return (st_A, st_B, st_R)

snd(st, N, M)
u||K||s_u||r_u ← st; H ← u||s_u
(C_1, C_2) ← enc_K(N, H, M);
s_u ← s_u + 1; st ← u||K||s_u||r_u
return (st, H, C_1, C_2)

tag(st_R, id_s, H, C_2)
K_R ← st_R; u ← id_s
if u ∉ {A, B} :
   return (⊥, ⊥)
v ← {A, B} \ {u}
T_R ← tag_mac(K_R, u||v||H||C_2)
return (st_R, T_R)

rcv(st, N, H, C_1, C_2)
u||K||s_u||r_u ← st; v||s_v ← H
(M, K_f) ← dec_K(N, H, C_1, C_2)
if r_u ≠ s_v or (M, K_f) =⊥:
   return (⊥, ⊥)
r_u ← r_u + 1
st ← u||K||s_u||r_u
return (st, M, K_f)

rprt(st_R, id_r, H, M, K_f, C_2, T_R)
K_R ← st_R; u ← id_r
if u ∉ {A, B} :
   return (⊥, 0)
v ← {A, B} \ {u}
if vrfy_mac(K_R, v||u||H||C_2, T_R) = 0 :
   return (⊥, 0)
if vrfy_nCE(H, M, K_f, C_2) = 0 :
   return (⊥, 0)
return (st_R, 1)
```

Fig. 7. Instantiation of a real MFC.

6.2 Security analysis

Confidentiality. The ciphertexts sent by MF are output by the nonce-base committing AEAD nCE. Then, if an adversary can differentiate between real and random ciphertexts output by MF, then it can differentiate between real and random ciphertexts output by nCE. Therefore, if nCE is MO-nRoR secure, then MF is MO-2nRoR secure. We state this formally in Theorem 1, skipping the proof that simply follows from the observation given above.

Theorem 1 (MF confidentiality). *Let* MF *be the* MFC *given in Figure 7, based on a secure committing AEAD scheme* nCE*. Then, for any adversary $\mathcal{A}$ playing the MO-2nRoR game there exists an adversary $\mathcal{B}$ such that*

$$\mathsf{Adv}_{\mathsf{MF}}^{mo\text{-}2nror}(\mathcal{A}) \leq \mathsf{Adv}_{\mathsf{nCE}}^{mo\text{-}nror}(\mathcal{B})$$

Integrity. The integrity property of MF follows from the use of send and receive counters and from the integrity of nCE. Formally, the following theorem holds:

Theorem 2. *Let* MF *be the* MFC *given in Figure 7, based on a committing AEAD scheme* nCE*. Then, for any adversary $\mathcal{A}$ playing the MO-2nCTXT game there exists an adversary $\mathcal{B}$ such that*

$$\mathsf{Adv}_{\mathsf{MF}}^{mo\text{-}2nctxt}(\mathcal{A}) \leq \mathsf{Adv}_{\mathsf{nCE}}^{mo\text{-}nctxt}(\mathcal{B})$$

Sender binding. The sender binding property follows directly from the sender binding property of nCE. Formally, we state the following theorem:

Theorem 3. *Let* MF *be the* MFC *given in Figure 7, based on a committing AEAD scheme* nCE*. Then, for any adversary* $\mathcal{A}$ *playing the S-2BIND game there exists an adversary* $\mathcal{B}$ *such that*

$$\mathsf{Adv}^{s\text{-}2bind}_{\mathsf{MF}}(\mathcal{A}) \leq \mathsf{Adv}^{s\text{-}bind}_{\mathsf{nCE}}(\mathcal{B})$$

Receiver binding. The receiver binding property of our construction is a consequence of the receiver binding property of the committing AEAD scheme and of the security against forgery of the MAC scheme used. We state the following theorem:

Theorem 4. *Let* MF *be the* MFC *given in Figure 7, based on a committing AEAD scheme* nCE *and a MAC scheme* MAC*. Then, for any adversary* $\mathcal{A}$ *playing the S-2BIND game there exists an adversary* $\mathcal{B}$ *and an adversary* $\mathcal{C}$ *such that*

$$\mathsf{Adv}^{r\text{-}2bind}_{\mathsf{MF}}(\mathcal{A}) \leq \mathsf{Adv}^{r\text{-}bind}_{\mathsf{nCE}}(\mathcal{B}) + \mathsf{Adv}^{uf\text{-}cma}_{\mathsf{MAC}}(\mathcal{C})$$

Due to space limits the proofs for Theorems $2, 3, 4$ are given in Appendix D.

7 Conclusion

In this paper, we introduced a *message franking channel* (MFC) which apart from message confidentiality/integrity, resistance to message drops, out of order delivery and replay attacks guarantees sender and receiver binding: Namely, the sender cannot send an abusive message, which cannot be reported to a third party and the receiver cannot report a fake message. Even if all of the definitions were presented in the unidirectional and bidirectional case, we focused mainly on bidirectional security definitions. We presented two results on binding properties in the canonic composition of two unidirectional MFC. In particular, we proved that two unidirectional receiver binding MFC are sufficient to create a bidirectional receiver binding MFC. In addition, we stressed that these results do not necessarily hold in general but only in the canonic composition. Finally, we gave an instantiation of a bidirectional MFC given a secure nonce based committing AEAD [GLR17] and a message authentication code.

References

BDPS13. Alexandra Boldyreva, Jean Paul Degabriele, Kenneth G Paterson, and Martijn Stam. On symmetric encryption with distinguishable decryption failures. In *International Workshop on Fast Software Encryption*, pages 367–390. Springer, 2013.

BH17. Colin Boyd and Britta Hale. Secure channels and termination: The last word on tls. Cryptology ePrint Archive, Report 2017/784, 2017. https://eprint.iacr.org/2017/784.

BHMS15. Colin Boyd, Britta Hale, Stig Frode Mjolsnes, and Douglas Stebila. From stateless to stateful: Generic authentication and authenticated encryption constructions with application to tls. Cryptology ePrint Archive, Report 2015/1150, 2015. https://eprint.iacr.org/2015/1150.

BKN02. Mihir Bellare, Tadayoshi Kohno, and Chanathip Namprempre. Authenticated encryption in ssh: provably fixing the ssh binary packet protocol. In *Proceedings of the 9th ACM conference on Computer and communications security*, pages 1–11. ACM, 2002.

DGRW18. Yevgeniy Dodis, Paul Grubbs, Thomas Ristenpart, and Joanne Woodage. Fast message franking: From invisible salamanders to encryptment. In *Advances in Cryptology - CRYPTO 2018 - 38th Annual International Cryptology Conference, Santa Barbara, CA, USA*, 2018.

Fac16. Facebook. Messenger secret conversations, 2016. https://fbnewsroomus.files.wordpress.com/2016/07/secret_conversations_whitepaper-1.pdf.

FMB+14. Tilman Frosch, Christian Mainka, Christoph Bader, Florian Bergsma, Joerg Schwenk, and Thorsten Holz. How secure is textsecure? Cryptology ePrint Archive, Report 2014/904, 2014. https://eprint.iacr.org/2014/904.

GLR17. Paul Grubbs, Jiahui Lu, and Thomas Ristenpart. Message franking via committing authenticated encryption. Cryptology ePrint Archive, Report 2017/664, 2017. https://eprint.iacr.org/2017/664.

GM17. Felix Günther and Sogol Mazaheri. A formal treatment of multi-key channels. Cryptology ePrint Archive, Report 2017/501, 2017. https://eprint.iacr.org/2017/501.

KPB03. Tadayoshi Kohno, Adriana Palacio, and John Black. Building secure cryptographic transforms, or how to encrypt and mac. Cryptology ePrint Archive, Report 2003/177, 2003. https://eprint.iacr.org/2003/177.

Mar17. Giorgia Azzurra Marson. *Real-World Aspects of Secure Channels: Fragmentation, Causality, and Forward Security*. PhD thesis, Technische Universität Darmstadt, 2017.

MP17. Giorgia Azzurra Marson and Bertram Poettering. Security notions for bidirectional channels. Cryptology ePrint Archive, Report 2017/161, 2017. https://eprint.iacr.org/2017/161.

Per18. Trevor Perrin. The noise protocol framework. http://noiseprotocol.org/noise.html, 2018. Accessed: 2018-09-03.

PR18. Bertram Poettering and Paul Rösler. Ratcheted key exchange, revisited. Cryptology ePrint Archive, Report 2018/296, 2018. https://eprint.iacr.org/2018/296.

Vau02. Serge Vaudenay. Security flaws induced by cbc padding — applications to ssl, ipsec, wtls... In Lars R. Knudsen, editor, *Advances in Cryptology — EUROCRYPT 2002*, pages 534–545, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg.

Wha17. Whatsapp. Whatsapp encryption overview, 2017. https://www.whatsapp.com/security/WhatsApp-Security-Whitepaper.pdf.

A Security of committing AEAD

Multi and single opening security. In some cases, one would like to encrypt multiple ciphertexts with the same key and keep all ciphertexts secure even if an opening is known. This notion is called multi-opening security. For example, if the key used for encryption is the same as the opening key (i.e. $K = K_f$), the scheme would not be multi-opening secure since the knowledge of one opening K_f would compromise all ciphertexts encrypted with the key K. On the other hand, if the secret key is meant to be used only once (e.g. Signal protocol), we can require only single-opening security. In this paper, we focus solely on multi-opening secure schemes.
Confidentiality. The multiple-opening confidentality of nonce-based committing AEAD is defined in MO-nREAL and MO-nRAND games (cf. 8). In these games an adversary $\mathcal{A}$ tries to differentiate between ciphertexts (C_1, C_2) derived from a legit message and ciphertexts derived from random bitstrings.

The nonce-based multiple opening real-or-random (MO-nROR) advantage of any adversary $\mathcal{A}$ as

$$\mathsf{Adv}^{\text{mo-nror}}_{\mathsf{nCE}}(\mathcal{A}) = |\Pr[\text{MO-nREAL}_{\mathsf{nCE}}(\mathcal{A}) \Rightarrow 1] - \Pr[\text{MO-nRAND}_{\mathsf{nCE}}(\mathcal{A}) \Rightarrow 1]|$$

Integrity. The multiple-opening integrity of nonce-based committing AEAD is defined with the MO-nCTXT game (cf. Figure 9). In this game, an adversary can query an encryption oracle $\mathcal{O}^{enc}$ and a decryption oracle $\mathcal{O}^{dec}$. Then, it sends a header, a nonce and ciphertexts (N, H, C_1, C_2) to a challenge oracle $\mathcal{O}^{chal}$. If the decryption is successful and the input values were not previously output by $\mathcal{O}^{enc}$, the adversary wins. Informally, the adversary wins if it can forge a valid tuple (N, H, C_1, C_2). We define the nonce-based multiple opening ciphertext integrity (MO-nCTXT) advantage of any adversary $\mathcal{A}$ playing the MO-nCTXT game as

$$\mathsf{Adv}^{\text{mo-nctxt}}_{\mathsf{nCE}}(\mathcal{A}) = \Pr[\text{MO-nCTXT}_{\mathsf{nCE}}(\mathcal{A}) \Rightarrow \mathsf{true}]$$

Binding security. In addition to the standard security definitions (i.e. confidentiality and integrity), Grubbs et al. [GLR17] defined two binding security notions, namely *sender binding* and *receiver binding.*

The intuition behind sender binding is that a participant cannot send a message which cannot be reported later. The s-BIND game in Figure 10 defines this security notion. An adversary wins if it can find a ciphertext and a franking tag such that a receiving participant can decrypt but the verification of the franking tag with the obtained message fails. For any adversary $\mathcal{A}$ playing s-BIND and any committing AEAD nCE, we define the advantage of $\mathcal{A}$ as

$$\mathsf{Adv}^{\text{s-bind}}_{\mathsf{nCE}}(\mathcal{A}) = \Pr[\text{s-BIND}_{\mathsf{CE}}(\mathcal{A}) \Rightarrow \mathsf{true}]$$

The idea of receiver binding is that a malicious receiver should not be able to accuse another client of sending an abusive message that was never sent (i.e. it should not be able to report a message that was not sent). The game r-BIND

MO-nREAL$_{\mathsf{nCE}}(\mathcal{A})$	MO-nRAND$_{\mathsf{nCE}}(\mathcal{A})$
$K \leftarrow\$ \mathsf{init}$ $E \leftarrow []; i \leftarrow 0$ $b' \leftarrow \mathcal{A}^{\mathcal{O}^{enc},\mathcal{O}^{dec},\mathcal{O}^{chal}}$ **return** b'	$K \leftarrow\$ \mathsf{init}$ $E \leftarrow []; i \leftarrow 0$ $b' \leftarrow \mathcal{A}^{\mathcal{O}^{enc},\mathcal{O}^{dec},\mathcal{O}^{chal}}$ **return** b'
Oracle $\mathcal{O}^{enc}(N, H, M)$ $(C_1, C_2) \leftarrow \mathsf{enc}_K(N, H, M)$ $E[i] \leftarrow (N, H, C_1, C_2)$ $i \leftarrow i + 1$ **return** (C_1, C_2)	**Oracle** $\mathcal{O}^{enc}(N, H, M)$ $(C_1, C_2) \leftarrow \mathsf{enc}_K(N, H, M)$ $E[i] \leftarrow (N, H, C_1, C_2)$ $i \leftarrow i + 1$ **return** (C_1, C_2)

Oracle $\mathcal{O}^{dec}(N, H, C_1, C_2)$
if (N, H, C_1, C_2) not in $E[]$:
  return $\perp$
$(M, K_f) \leftarrow \mathsf{dec}_K(N, H, C_1, C_2)$
return (M, K_f)
	Oracle $\mathcal{O}^{dec}(N, H, C_1, C_2)$
	if (N, H, C_1, C_2) not in $E[]$:
	  return $\perp$
	$(M, K_f) \leftarrow \mathsf{dec}_K(N, H, C_1, C_2)$
	return (M, K_f)

Oracle $\mathcal{O}^{chal}(N, H, M)$ $(C_1, C_2) \leftarrow \mathsf{enc}_K(N, H, M)$ **return** (C_1, C_2)	**Oracle** $\mathcal{O}^{chal}(N, H, M)$ $H_r \leftarrow\$ \{0,1\}^{\lvert H \rvert}$ $M_r \leftarrow\$ \{0,1\}^{\lvert M \rvert}$ $(C_1, C_2) \leftarrow \mathsf{enc}_K(N, H_r, M_r)$ **return** (C_1, C_2)

Fig. 8. Confidentality games for nonce-based committing AEAD

in Figure 10 defines this security notion. An adversary wins if it can find two different messages such that their commitments are the same. For any adversary $\mathcal{A}$ playing the r-BIND game and any committing AEAD nCE, we define the advantage of $\mathcal{A}$ as

$$\mathsf{Adv}^{\text{r-bind}}_{\mathsf{nCE}}(\mathcal{A}) = \Pr[\text{r-BIND}_{\mathsf{CE}}(\mathcal{A}) \Rightarrow \mathsf{true}]$$

B Unidirectional definitions

B.1 Correctness

Unidirectional correctness. For a unidirectional channel (i.e. only Alice sends messages to Bob), the game of correctness is equivalent to the bidirectional game of correctness as we see in Figure 11. The main differences are that the participants are fixed in each oracle, the sender is always clean (since it cannot receive messages) and we need less variables (only one copy of s, r, h, M and C).

$\text{MO-nCTXT}_{\mathsf{nCE}}(\mathcal{A})$

$K \leftarrow\$ \mathsf{init}; win \leftarrow \mathsf{false}$
$E \leftarrow []; i \leftarrow 0$
$\mathcal{A}^{\mathcal{O}^{enc}, \mathcal{O}^{dec}, \mathcal{O}^{chal}}$
return win

Oracle $\mathcal{O}^{enc}(N, H, M)$

$(C_1, C_2) \leftarrow \mathsf{enc}_K(N, H, M)$
$E[i] \leftarrow (N, H, C_1, C_2)$
$i \leftarrow i + 1$
return (C_1, C_2)

Oracle $\mathcal{O}^{dec}(N, H, C_1, C_2)$

return $\mathsf{dec}_K(N, H, C_1, C_2)$

Oracle $\mathcal{O}^{chal}(N, H, C_1, C_2)$

if (N, H, C_1, C_2) in $E[]$:
  return $\perp$
$(M, K_f) \leftarrow \mathsf{dec}_K(N, H, C_1, C_2)$
if $(M, K_f) \neq \perp$:
  $win \leftarrow \mathsf{true}$
return (M, K_f)

Fig. 9. Integrity game for nonce-based committing AEAD

$\text{r-BIND}_{\mathsf{nCE}}(\mathcal{A})$

$win \leftarrow \mathsf{false}$
$(H, M, K_f, H', M', K'_f, C_2) \leftarrow\$ \mathcal{A}$
$b \leftarrow \mathsf{vrf}(H, M, K_f, C_2)$
$b' \leftarrow \mathsf{vrf}(H', M', K'_f, C_2)$
if $M \neq M'$ and $b = b' = 1$:
  $win \leftarrow \mathsf{true}$
return win

$\text{s-BIND}_{\mathsf{nCE}}(\mathcal{A})$

$win \leftarrow \mathsf{false}$
$(N, H, C_1, C_2, K) \leftarrow\$ \mathcal{A}$
$(M, K_f) \leftarrow \mathsf{dec}_K(N, H, C_1, C_2)$
$b \leftarrow \mathsf{vrf}(H, M, K_f, C_2)$
if $(M, K_f) \neq \perp$ and $b = 0$:
  $win \leftarrow \mathsf{true}$
return win

Fig. 10. Games for committing AEAD binding notions.

B.2 Binding Security

We present in Figure 12 the games r-1BIND and s-1BIND defining unidirectional receiver and sender binding notions, respectively. These games are similar to their bidirectional equivalent. However, in a unidirectional MFC, if A is the sender and B is the receiver, then $\mathsf{snd}(st_u, *)$ fails if $u \neq A$, $\mathsf{tag}(st_R, id_s, *)$ fails if $id_s \neq A$, $\mathsf{rprt}(st_R, id_r, *)$ fails if $id_r \neq B$ and $\mathsf{rcv}(st_u, *)$ fails if $u \neq B$. Therefore, the adversary can query $\mathcal{O}^{snd}, \mathcal{O}^{tag}$ only with $u = A$ and $\mathcal{O}^{rcv}, \mathcal{O}^{rprt}$ only with $u = B$. For simplicity we omit the argument u in all oracles.

C Binding notions in the canonic composition

A bidirectional standard cryptographic channel made of two secure unidirectional channels is not necessarily secure [MP17]. Therefore, it is of interest to study the binding security of such constructions adapted to the MFC. More precisely, we want to analyze the relations between unidirectional and bidirectional receiver/sender binding, in a special channel called the *canonic composition* [MP17]. This channel captures the idea that real-life communication proto-

```
1-CORR_Ch(A)
win ← false
(st_A, st_B, st_R) ←$ init
s ← 0; r ← 0; h ← true
M ← []; C ← []
A^{O^snd, O^rcv, O^rprt}
return win

Oracle O^rprt(n, H, M, K_f, C_2, T_R)
st_B ← rprt(st_B, N)
(st_R, b) ← rprt(st_R, B, H, M, K_f, C_2, T_R)
if h :
  if C[n] = (*, H, *, C_2, T_R) and M[n] = M :
    if b = 0 :
      win ← true
  else :
    h ← false
return b

Oracle O^snd(N, M)
(st_A, H, C_1, C_2) ← snd(st_A, N, H, M)
(st_R, T_R) ← tag(st_R, A, H, C_2)
M[s] ← M
C[s] ← (N, H, C_1, C_2, T_R)
s ← s + 1
return (C_1, C_2, T_R)

Oracle O^rcv(N, H, C_1, C_2)
(st_B, M, K_f) ← rcv(st_B, N, H, C_1, C_2)
if h :
  if r < s and C[r] = (N, H, C_1, C_2, *) :
    if (M, K_f) =⊥ or M[r] ≠ M :
      win ← true
    r ← r + 1
  else :
    h ← false
return (M, K_f)
```

Fig. 11. Correctness game for a unidirectional message franking channel.

cols are often designed to be the composition of two independent unidirectional secure channels.

C.1 Message franking canonic composition

The canonic construction is the straightforward composition of two unidirectional MFC, when one party sends and the other receives and reports. Let $\mathsf{Ch} = (\mathsf{init}, \mathsf{snd}, \mathsf{tag}, \mathsf{rcv}, \mathsf{rprt})$ be a unidirectional MFC with a participants space $\mathcal{P}$, a key space $\mathcal{K}$, a nonce space $\mathcal{N}$, a header space $\mathcal{H}$, a message space $\mathcal{M}$, a ciphertext space $\mathcal{C}$, an opening key space $\mathcal{K}_f$, a franking tag space $\mathcal{T}$, a router tag space $\mathcal{T}_R$ and a state space $\mathcal{S}$. The canonic composition will use one instance of Ch for the communication from Alice to Bob ($\rightarrow$) and another for the communication from Bob to Alice ($\leftarrow$).
Let $\mathsf{Ch}' = (\mathsf{init}', \mathsf{snd}', \mathsf{tag}', \mathsf{rcv}', \mathsf{rprt}')$ be the bidirectional MFC resulting from the canonical composition of Ch described in Figure 13. Associated to this channel Ch' is a participants space $\mathcal{P}' = \mathcal{P}$, a key space $\mathcal{K}' = \mathcal{K} \times \mathcal{K}$, a nonce space $\mathcal{N}' = \mathcal{N}$, a header space $\mathcal{H}' = \mathcal{H}$, a message space $\mathcal{M}' = \mathcal{M}$, a ciphertext space $\mathcal{C}' = \mathcal{C}$, an opening key space $\mathcal{K}'_f = \mathcal{K}_f$, a franking tag space $\mathcal{T}' = \mathcal{T}$, a router tag space $\mathcal{T}'_R = \mathcal{T}_R$ and a state space $\mathcal{S}' = \mathcal{S} \times \mathcal{S}$.

A participant's state consists of a state used to receive and another to send, initialized in the init' procedure. This procedure also creates a router state for the $\rightarrow$ channel and another for the $\leftarrow$ channel. Then, when a party wants to

r-1BIND$_{\mathsf{Ch}}(\mathcal{A})$

$win \leftarrow \mathsf{false}$
$(st_A, st_B, st_R) \leftarrow\$ \mathsf{init}$
$s \leftarrow 0; M \leftarrow []$
$\mathcal{A}^{\mathcal{O}^{snd}, \mathcal{O}^{rprt}, st_A, st_B}$
return win

Oracle $\mathcal{O}^{snd}(N, H, M)$

$id_s \leftarrow A$
$(st_u, _, C_1, C_2) \leftarrow \mathsf{snd}(st_u, N, M)$
$(st_R, T_R) \leftarrow \mathsf{tag}(st_R, id_s, H, C_2)$
$M[s] \leftarrow M$
$s \leftarrow s + 1$
return (C_1, C_2, T_R)

Oracle $\mathcal{O}^{rprt}(H, M, K_f, C_2, T_R)$

$id_r \leftarrow B$
$(st_R, b) \leftarrow \mathsf{rprt}(st_R, id_r, H, M, K_f, C_2, T_R)$
if M **not** in $M[]$ and $b = 1$:
  $win \leftarrow \mathsf{true}$
return b

s-1BIND$_{\mathsf{Ch}}(\mathcal{A})$

$win \leftarrow \mathsf{false}$
$(st_A, st_B) \leftarrow\$ \mathsf{init}$
$st_R \leftarrow\$ \mathsf{init}_R$
$s \leftarrow 0; r \leftarrow 0$
$S \leftarrow []; R \leftarrow []$
$\mathcal{A}^{\mathcal{O}^{tag}, \mathcal{O}^{rcv}, \mathcal{O}^{rprt}, st_A, st_B}$
return win

Oracle $\mathcal{O}^{tag}(H, C_2)$

$id_s \leftarrow A$
$(st_R, T_R) \leftarrow \mathsf{tag}(st_R, id_s, H, C_2)$
$S[s] = (H, C_2, T_R)$
$s \leftarrow s + 1$
return T_R

Oracle $\mathcal{O}^{rcv}(N, H, C_1, C_2)$

if (H, C_2, T_R) in S for some T_R :
  $(st_B, M, K_f) \leftarrow rcv(st_B, N, H, C_1, C_2)$
  if $(M, K_f) \neq \perp$
    $R[r] \leftarrow (H, M, K_f, C_2, T_R)$
    $r \leftarrow r + 1$

Oracle $\mathcal{O}^{rprt}(H, M, K_f, C_2, T_R)$

$id_r \leftarrow B$
$(st_R, b) \leftarrow \mathsf{rprt}(st_R, id_r, H, M, K_f, C_2, T_R)$
if (H, M, K_f, C_2, T_R) in R and $b = 0$:
  $win \leftarrow \mathsf{true}$
return b

Fig. 12. Games for receiver and sender binding security definitions in the unidirectional MFC case.

send a message, snd' extracts the sending state and uses it as the state in the snd call. The sender and reporter's identities id_s and id_r are used in tag' and rprt' to determine the direction of the communication (i.e. $\rightarrow$ or $\leftarrow$). This is necessary for choosing the right router state st_R, which is itself needed to invoke tag and rprt. When a client receives a message, the receiving state is extracted and rcv is called. If an error occurs during this procedure, the state is set to $\perp$.

C.2 Security analysis

Let Ch be a unidirectional MFC and Ch' be the bidirectional MFC resulting from the canonic composition of two channels Ch. We study the relations between

init'

$(st_{snd}^{\rightarrow}, st_{rcv}^{\rightarrow}, st_R^{\rightarrow}) \leftarrow\$ \mathsf{init}$
$(st_{snd}^{\leftarrow}, st_{rcv}^{\leftarrow}, st_R^{\leftarrow}) \leftarrow\$ \mathsf{init}$
$st_A \leftarrow (st_{snd}^{\rightarrow}, st_{rcv}^{\leftarrow})$
$st_B \leftarrow (st_{snd}^{\leftarrow}, st_{rcv}^{\rightarrow})$
$st_R \leftarrow (st_R^{\rightarrow}, st_R^{\leftarrow})$
return (st_A, st_B, st_R)

$\mathsf{snd}'(st, N, M)$

$(st_{snd}, st_{rcv}) \leftarrow st$
$(st_{snd}, H, C_1, C_2) \leftarrow \mathsf{snd}(st_{snd}, N, M)$
$st \leftarrow (st_{snd}, st_{rcv})$
return (st, H, C_1, C_2)

$\mathsf{tag}'(st_R, id_s, H, C_2)$

$(st_R^{\rightarrow}, st_R^{\leftarrow}) \leftarrow st_R$
if $id_s = A$:
  $(st_R^{\rightarrow}, T_R) \leftarrow \mathsf{tag}(st_R^{\rightarrow}, H, C_2)$
else :
  $(st_R^{\leftarrow}, T_R) \leftarrow \mathsf{tag}(st_R^{\leftarrow}, H, C_2)$
$st_R \leftarrow (st_R^{\rightarrow}, st_R^{\leftarrow})$
return (st_R, T_R)

$\mathsf{rcv}'(st, N, H, C_1, C_2)$

$(st_{snd}, st_{rcv}) \leftarrow st$
$(st_{rcv}, M, K_f) \leftarrow \mathsf{rcv}(st_{rcv}, N, H, C_1, C_2)$
if $(M, K_f) \neq \perp$:
  $st \leftarrow (st_{snd}, st_{rcv})$
else :
  $st \leftarrow \perp$
return (st, M, K_f)

$\mathsf{rprt}'(st_R, id_r, H, M, K_f, C_2, T_R)$

$(st_R^{\rightarrow}, st_R^{\leftarrow}) \leftarrow st_R$
if $id_r = B$:
  $(st_R^{\rightarrow}, b) \leftarrow \mathsf{rprt}(st_R^{\rightarrow}, H, M, K_f, C_2, T_R)$
else :
  $(st_R^{\leftarrow}, b) \leftarrow \mathsf{rprt}(st_R^{\leftarrow}, H, M, K_f, C_2, T_R)$
$st_R \leftarrow (st_R^{\rightarrow}, st_R^{\leftarrow})$
if $b = 1$: **return** $(st_R, 1)$
else : **return** $(\perp, 0)$

Fig. 13. Canonic composition of two unidirectional MFC.

the bidirectional binding notions r-2BIND, s-2BIND and their equivalent in the unidirectional case r-1BIND, s-1BIND. We prove the two following results:

$$\text{r-1BIND} + \text{r-1BIND} \iff \text{r-2BIND} \tag{4}$$
$$\text{s-1BIND} + \text{s-1BIND} \iff \text{s-2BIND} \tag{5}$$

The $\Leftarrow$ direction is trivial in both cases, as if one is able to break the binding security of one of the two channels composing Ch', then Ch' will not be binding either. Therefore, we state only the $\Longrightarrow$ relations.

Theorem 5 (Receiver binding). *Let* Ch *be a receiver binding unidirectional* MFC *(r-1BIND). Then, its canonic composition* Ch' *is receiver binding (r-2BIND). More precisely, for any adversary* $\mathcal{A}$ *playing r-2BIND against* Ch'*, there exists a r-1BIND adversary* $\mathcal{B}$ *against* Ch *s.t.*

$$\mathsf{Adv}^{r\text{-}2bind}_{\mathsf{Ch}'}(\mathcal{A}) \leq 2 \cdot \mathsf{Adv}^{r\text{-}1bind}_{\mathsf{Ch}}(\mathcal{B})$$

Proof. In order to prove this statement, we follow the same method used to prove Theorem 2 in [MP17]. Let $win^{\rightarrow}$ be the event that $\mathcal{A}$ wins the r-2BIND game by querying $\mathcal{O}^{rprt}(u, *)$ with $u = B$. In other words, $win^{\rightarrow}$ happens when line 4 of $\mathcal{O}^{rprt}$ in Figure 5 is executed and $u = B$. Similarly, $win^{\leftarrow}$ is the event that line 4 of $\mathcal{O}^{rprt}$ in Figure 5 is executed and $u = A$.
Now, we consider a sequence of games. Let G^0 be the r-2BIND game played by adversary $\mathcal{A}$ against the channel Ch'. Let G^1 be the same game as G^0 but the game returns false when $win^{\rightarrow}$ happens. Let G^2 be the same as G^1 but the game returns false when $win^{\leftarrow}$ happens. We deduce the following results:

$$|\Pr[G^0 \Rightarrow \mathsf{true}] - \Pr[G^1 \Rightarrow \mathsf{true}]| \leq \Pr[win^{\rightarrow}]$$
$$|\Pr[G^1 \Rightarrow \mathsf{true}] - \Pr[G^2 \Rightarrow \mathsf{true}]| \leq \Pr[win^{\leftarrow}]$$
$$\Pr[G^2 \Rightarrow \mathsf{true}] = 0$$

Also,

$$\begin{aligned}\mathsf{Adv}^{\text{r-2bind}}_{\mathsf{Ch}'}(\mathcal{A}) &= \Pr[G^0 \Rightarrow \mathsf{true}] \\ &\leq |\Pr[G^0 \Rightarrow \mathsf{true}] - \Pr[G^1 \Rightarrow \mathsf{true}]| \\ &+ |\Pr[G^2 \Rightarrow \mathsf{true}] - \Pr[G^1 \Rightarrow \mathsf{true}]| \\ &+ \Pr[G^2 \Rightarrow \mathsf{true}]\end{aligned}$$

Hence,

$$\mathsf{Adv}^{\text{r-2bind}}_{\mathsf{Ch}'}(\mathcal{A}) \leq \Pr[win^{\rightarrow}] + \Pr[win^{\leftarrow}] \tag{6}$$

Now, given the adversary $\mathcal{A}$, we construct a r-1BIND adversary $\mathcal{B}^{\rightarrow}$ (respectively $\mathcal{B}^{\leftarrow}$) against channel Ch that wins with probability $\Pr[win^{\rightarrow}]$ (respectively

$\Pr[win^{\leftarrow}]$). Consider the adversary $\mathcal{B}^{\rightarrow}$ playing the r-1BIND game with channel Ch in Figure 14. It has access to the $\mathcal{O}^{snd}$ and $\mathcal{O}^{rprt}$ oracles and the states $st^{\rightarrow}_{snd}, st^{\rightarrow}_{rcv}$ (i.e. st_A and st_B in the r-1BIND game in Figure 12). The adversary $\mathcal{B}^{\rightarrow}$ creates its own instance of Ch by sampling the necessary states randomly (line 1-4). Then, $\mathcal{B}^{\rightarrow}$ runs $\mathcal{A}$. The latter calls $\mathcal{O}^{snd}_{\mathcal{A}}$ and $\mathcal{O}^{rprt}_{\mathcal{A}}$ oracles. In both of these oracles, $\mathcal{A}$'s queries in the $\rightarrow$ direction (Alice sends/Bob reports) are relayed to $B^{\rightarrow}$'s own oracles (line 3 in $\mathcal{O}^{snd}_{\mathcal{A}}$ and $\mathcal{O}^{rprt}_{\mathcal{A}}$). $\mathcal{A}$'s queries in the other direction are answered by $B^{\rightarrow}$ running its own channel (lines 5-8 in $\mathcal{O}^{snd}_{\mathcal{A}}$ and line 5 in $\mathcal{O}^{rprt}_{\mathcal{A}}$). Overall, $B^{\rightarrow}$ creates a canonic composition of its own channel for the $\leftarrow$ direction and of the channel created in the r-1BIND game for the $\rightarrow$ direction. $\mathcal{A}$ runs exactly as in the G^0 game, except when it wins. Now, consider $\mathcal{A}$ winning in the G^0 game with a $\rightarrow$ report query (i.e. $win^{\rightarrow}$ event). Then, in the r-1BIND game, the winning query is forwarded to $\mathcal{O}^{rprt}$ by $\mathcal{B}^{\rightarrow}$ and line 5 of $\mathcal{O}^{rprt}$ in the r-1BIND game of Figure 12 is executed, making $\mathcal{B}^{\rightarrow}$ win. $\mathcal{B}^{\rightarrow}$ wins because the array M in the r-1BIND game played by $\mathcal{B}^{\rightarrow}$ is modified exactly as the array $M_{A,B}$ in the game G^0 played by $\mathcal{A}$. Thus, if $win^{\rightarrow}$ happens, then the condition allowing $\mathcal{B}^{\rightarrow}$ to win in r-1BIND (line 4 of $\mathcal{O}^{rprt}$ in Figure 12) evaluates to true. Therefore, we obtain

$$\mathsf{Adv}^{\text{r-1bind}}_{\mathsf{Ch}}(\mathcal{B}^{\rightarrow}) \geq \Pr[win^{\rightarrow}]$$

Similarly, one can construct an adversary $\mathcal{B}^{\leftarrow}$ by inverting the directions accordingly in Figure 14, obtaining

$$\mathsf{Adv}^{\text{r-1bind}}_{\mathsf{Ch}}(\mathcal{B}^{\leftarrow}) \geq \Pr[win^{\leftarrow}]$$

Finally, we construct an adversary $\mathcal{B}$ that chooses uniformly at random to run $\mathcal{B}^{\rightarrow}$ or $\mathcal{B}^{\leftarrow}$, achieving a total advantage of

$$\mathsf{Adv}^{\text{r-1bind}}_{\mathsf{Ch}}(\mathcal{B}) \geq \frac{1}{2}\Pr[win^{\rightarrow}] + \frac{1}{2}\Pr[win^{\leftarrow}]$$

Hence, by Eq. (6), we conclude that

$$\mathsf{Adv}^{\text{r-2bind}}_{\mathsf{Ch}'}(\mathcal{A}) \leq \Pr[win^{\rightarrow}] + \Pr[win^{\leftarrow}] \leq 2 \cdot \mathsf{Adv}^{\text{r-1bind}}_{\mathsf{Ch}}(\mathcal{B})$$

□

Theorem 6 (Sender binding). *Let* Ch *be a sender binding unidirectional* MFC *(r-1BIND). Then, its canonic composition* Ch' *is sender binding (s-2BIND). More precisely, for any adversary* $\mathcal{A}$ *playing s-2BIND against* Ch'*, there exists a s-1BIND adversary* $\mathcal{B}$ *against* Ch *s.t.*

$$\mathsf{Adv}^{s\text{-}2bind}_{\mathsf{Ch}'}(\mathcal{A}) \leq 2 \cdot \mathsf{Adv}^{s\text{-}1bind}_{\mathsf{Ch}}(\mathcal{B})$$

Proof. The proof is similar to the one of Theorem 5.

$\mathcal{B}^{\rightarrow}(\mathcal{O}^{snd}, \mathcal{O}^{rprt}, st^{\rightarrow}_{snd}, st^{\rightarrow}_{rcv})$

1 : $(st^{\leftarrow}_{snd}, st^{\leftarrow}_{rcv}, st^{\leftarrow}_{R}) \leftarrow\$ \mathsf{init}$
2 : $st_A \leftarrow (st^{\rightarrow}_{snd}, st^{\leftarrow}_{rcv})$
3 : $st_B \leftarrow (st^{\leftarrow}_{snd}, st^{\rightarrow}_{rcv})$
4 : $\mathcal{A}^{\mathcal{O}^{tag}_{\mathcal{A}}, \mathcal{O}^{rprt}_{\mathcal{A}}, st_A, st_B}$

Oracle $\mathcal{O}^{rprt}_{\mathcal{A}}(u, H, M, K_f, C_2, T_R)$

1 : $id_r \leftarrow u$
2 : **if** $u \neq A$:
3 : $\quad b \leftarrow \mathcal{O}^{rprt}(H, M, K_f, C_2, T_R)$
4 : **else** :
5 : $\quad (st^{\leftarrow}_{R}, b) \leftarrow \mathsf{rprt}(st^{\leftarrow}_{R}, id_r, H, M, K_f, C_2, T_R)$
6 : **return** b

Oracle $\mathcal{O}^{snd}_{\mathcal{A}}(u, N, H, M)$

1 : $id_s = u$
2 : **if** $u \neq B$:
3 : $\quad (C_1, C_2, T_R) \leftarrow \mathcal{O}^{snd}(N, H, M)$
4 : **else** :
5 : $\quad (st_{snd}, st_{rcv}) \leftarrow st_B$
6 : $\quad (st_{snd}, _, C_1, C_2) \leftarrow \mathsf{snd}(st_{snd}, N, M)$
7 : $\quad (st^{\leftarrow}_{R}, T_R) \leftarrow \mathsf{tag}(st^{\leftarrow}_{R}, id_s, H, C_2)$
8 : $\quad st_B \leftarrow (st_{snd}, st_{rcv})$
9 : **return** C_1, C_2, T_R

Fig. 14. Reduction for the proof of Theorem 5.

C.3 Generic Composition

Both Theorem 1 and Theorem 2 hold in the case of the canonic composition but not for any composition. For example, one can easily design a bidirectional MFC which is not r-2BIND from two r-1BIND unidirectional MFC. Consider the canonic composition of Figure 13 but we modify it such that the communication keys and the router keys are the same in both directions (i.e. $K^{\rightarrow} = K^{\leftarrow}$ and $K^{\rightarrow}_{R} = K^{\leftarrow}_{R}$). That can be achieved by making only one call to init in init' and making each state equal in both directions. These modifications do not change the r-1BIND security of the channel. In particular, if the unidirectional MFC used in the composition is r-1BIND, then the composition is still r-1BIND in each direction.

Now consider the adversary $\mathcal{A}^{\text{r-2BIND}}$ of Figure 6 presented in Section 5.3. It is easy to see that this adversary will win with probability 1 against the new composition. Indeed, it is interesting to observe that the original canonic composition is r-2BIND secure because the router keys $K^{\rightarrow}_{R}$ and $K^{\leftarrow}_{R}$ are likely to be different. Therefore, they prevent $\mathcal{A}^{\text{r-2BIND}}$ from winning since the router key from the $\rightarrow$ channel is used to generate the router tag T_R but the adversary tries to report a message in the $\leftarrow$ channel. In fact, the router keys act as the identifiers of the sender/receiver in the router tag. Hence, if the router keys are the same for both directions (as in the modified composition), the resulting bidirectional MFC will not be r-2BIND secure.

D Proofs

Proof (Theorem 2). We give here a sketch of the proof. We consider the MO-2nCTXT game of Figure 3. For $\mathcal{A}$ to win either $s_v \leq r_u$ or $(N, H, C_1, C_2) \neq C_{v,u}[r_u]$ in $\mathcal{O}^{rcv}$. Bboth cannot happen since $C_{v,u}[r_u]$ is not defined if $r_u \geq s_v$.

Let $cond^1$ be the event that the adversary wins with the first condition evaluating to true. It means that MF decrypted successfully but more messages were received by u than sent by v. However, MF keeps track of the number of messages sent and received by putting the number of sent messages in the header. Therefore, $cond^1$ happens only if the rcv procedure of Figure 7 is invoked with $H' = v||s'_v$, where s'_v is greater or equal than the actual sent messages, and the decryption was successful. Therefore, we know that the adversary found (N', H', C'_1, C'_2) that decrypts successfully, with H' different from all headers previously sent.

Now, let $cond^2$ be the event that the adversary wins and the second condition is satisfied, which implies that $cond^1$ is false. As for the first condition, one can deduce by inspection that if $cond^2$ is true, then the adversary found a tuple (N', H', C'_1, C'_2) that was not output by the $\mathcal{O}^{snd}$ oracle before. Hence, these two arguments mean that one can construct an adversary $\mathcal{B}$ playing the MO-nCTXT game with nCE, which wins with a probability at least $\Pr[\mathcal{A} \text{ wins }] = \Pr[\{cond^1\} \cup \{cond^2\}]$.

Proof (Theorem 3). In the s-2BIND game (Figure 4), the adversary wins if the values (H, M, K_f, C_2, T_R) were successfully received by user u (i.e. it is in the $R_{v,u}$ array) but the rprt procedure fails. In the MF construction, rprt can fail for two reasons:

1. The first reason is that the verification of the router tag T_R failed. Now, in the s-2BIND game, if the adversary wins by querying $\mathcal{O}^{rprt}$ as client u, then the tuple (H, C_2, T_R) is in the $R_{v,u}$ array. This means that this tuple was previously in the $S_{v,u}$ array and therefore, T_R was output by $\mathsf{tag}(\mathsf{st_R}, \mathsf{v}, \mathsf{H}, \mathsf{C_2})$, with $v = \{A, B\} \setminus \{u\}$. Thus, $T_R = \mathsf{tag_{mac}}(K_R, v||u||H||C_2)$ and the router tag verification algorithm outputs the result of $\mathsf{vrfy_{mac}}(K_R, v||u||H||C_2, T_R)$, which can only be a success if we assume the MAC scheme to be correct. Hence, if the adversary wins, that contradicts the failure of the router tag verification procedure.
2. The second reason is that the franking tag verification $\mathsf{vrfy_{nCE}}(H, M, K_f, C_2)$ fails. Since the input values are in $R_{v,u}$ (assuming $\mathcal{A}$ queried $\mathcal{O}^{rprt}$ as user u), it means that rcv was successful, implying that $\mathsf{dec}_K(N, H, C_1, C_2)$ was successful. This means that $\mathcal{A}$ found a tuple (N, H, C_1, C_2, K) such that the decryption is successful but the verification fails.

Hence, given an adversary $\mathcal{A}$ who plays s-2BIND, it is trivial to construct an adversary $\mathcal{B}$ that wins the s-BIND game with at least the same probability as $\mathcal{A}$. □

Proof (Theorem 4). We proceed with the game hopping technique. Let the game G^0 in Figure 15 be the r-2BIND game played by adversary $\mathcal{A}$ against MF, except that the rprt procedure has been made explicit in $\mathcal{O}^{rprt}$ and that the router tags T_R output by the $\mathcal{O}^{snd}$ oracle are stored in an array $T[]$ along with the corresponding (u, v, H, C_2) tuple.

Now, consider the modified game G^1 (modifications boxed in Figure 15), which aborts if the router tag verification is successful but T_R was not obtained from a tag query with (v, u, H, C_2) in the sending oracle $\mathcal{O}^{snd}$. Let this condition be $forge$. Then, we have $|\Pr[G^0 \Rightarrow 1] - \Pr[G^1 \Rightarrow 1]| \leq \Pr[forge]$. By inspection, we see that if $forge$ happens, it means that the adversary found a message $m' = v||u||H||C_2$ and a corresponding tag T_R, without having queried $\mathsf{tag}_{\mathsf{mac}}(K_R, m')$. Hence, given $\mathcal{A}$ one can construct a straightforward UF-CMA adversary $\mathcal{C}$ against MAC that achieves an advantage of $\Pr[forge]$.

Let G^2 be the game G^1 but $\mathcal{O}^{rprt}$ is modified such that it aborts if everything is successful, but the message M is not in $M_{v,u}$ (modifications in grey in Figure 15). In other words, we modify G^1 such that the adversary never wins. If this condition is fulfilled in line 11 of $\mathcal{O}^{rprt}$, we know that (v, u, H, C_2, T_R) was obtained from a query to $\mathcal{O}^{snd}$ with some message M' because line 7 of $\mathcal{O}^{rprt}$ was not executed. Therefore, M' is in $M_{v,u}$ and thus $M \neq M'$. If this condition is satisfied, it means that the adversary found a tuple $(H, M, M', K_f, K'_f, C_2)$ with $M \neq M'$ such that $\mathsf{vrf}_{\mathsf{nCE}}(H, M, K_f, C_2) = \mathsf{vrf}_{\mathsf{nCE}}(H, M', K'_f, C_2) = 1$. Hence, given the adversary $\mathcal{A}$, one can construct a straightforward r-BIND adversary $\mathcal{B}$ against nCE such that $|\Pr[G^1 \Rightarrow 1] - \Pr[G^2 \Rightarrow 1]| \leq \mathsf{Adv}^{\text{r-bind}}_{\mathsf{nCE}}(\mathcal{B})$. Finally, we have $\Pr[G^2 \Rightarrow 1] = 0$, which concludes the theorem. □

$G^0(\mathcal{A})$, $\boxed{G^1(\mathcal{A})}$, $G^2(\mathcal{A})$

$win \leftarrow \mathsf{false}$
$(st_A, st_B, st_R) \leftarrow\$ \mathsf{init}$
$s_A \leftarrow 0; s_B \leftarrow 0$
$M_{A,B} \leftarrow []; M_{B,A} \leftarrow []$
$\mathcal{A}^{\mathcal{O}^{snd}, \mathcal{O}^{rprt}, st_A, st_B}$
$t \leftarrow 0; T \leftarrow []$
return win

Oracle $\mathcal{O}^{snd}(u, N, H, M)$

$id_s \leftarrow u$
$(st_u, _, C_1, C_2) \leftarrow \mathsf{snd}(st_u, N, M)$
$(st_R, T_R) \leftarrow \mathsf{tag}(st_R, id_s, H, C_2)$
$v = \{A, B\} \setminus \{u\}$
$M_{u,v}[s_u] \leftarrow M$
$T[t] \leftarrow (u, v, H, C_2, T_R); t \leftarrow t + 1$
$s_u \leftarrow s_u + 1$
return (C_1, C_2, T_R)

Oracle $\mathcal{O}^{rprt}(u, H, M, K_f, C_2, T_R)$

$id_r \leftarrow u; v = \{A, B\} \setminus \{u\}$
$K_R \leftarrow st_R; u \leftarrow id_r$
if $u \notin \{A, B\}$:
  $st_R \leftarrow \perp$; **return** 0
if $\mathsf{vrfy}_{\mathsf{mac}}(K_R, v||u||H||C_2, T_R) = 0$:
  $st_R \leftarrow \perp$; **return** 0
$\boxed{\textbf{if } (v, u, H, C_2, T_R) \textbf{ not in } T : \textbf{abort}}$
if $\mathsf{vrfy}_{\mathsf{nCE}}(H, M, K_f, C_2) = 0$:
  $st_R \leftarrow \perp$; **return** 0
if M **not** in $M_{v,u}$:
  abort
  $win \leftarrow \mathsf{true}$
return 1

Fig. 15. Games for proof of Theorem 4.