Authentication Capacity of Adversarial Channels

Oliver Kosut
School of Electrical, Computer and Energy Engineering
Arizona State University
Tempe, AZ 85287
Email: okosut@asu.edu

Jörg Kliewer
Department of Electrical and Computer Engineering
New Jersey Institute of Technology
Newark, NJ 07102
Email: jkliewer@njit.edu

***Abstract*—Keyless authentication is considered in an adversarial point-to-point channel. Namely, a legitimate transmitter and receiver aim to communicate over a noisy channel that may or may not also contain an active adversary, capable of transmitting an arbitrary signal into the channel. If the adversary is not present, then the receiver must successfully decode the message with high probability; if it is present, then the receiver must either decode the message or detect the adversary's presence. Thus, whenever the receiver decodes, it can be certain that the decoded message is authentic. The exact authentication capacity is characterized for discrete-memoryless adversary channels, where the adversary is assumed to know the code but not the message. The authentication capacity is shown to be either zero or equal to the no-adversary capacity, depending on whether the channel satisfies a condition termed overwritability.**

I. Introduction

Communication systems are often required to be reliable even in adversarial environments, in which an attacker can enter the medium, and inject unwanted, intelligently crafted, malicious signals. To this end, coding strategies are necessary that can overcome and adapt to the presence of an adversary.

The classical information theory model that captures an adversary in a point-to-point communication system is the arbitrarily-varying channel (AVC) [1], in which the decoder receives a signal based on both the legitimate transmission as well as the adversary transmission. In this problem, the adversary is assumed be always present and able to transmit an arbitrary signal across the coding block (subject to various knowledge and power assumptions), and the goal is to recover the legitimate message no matter what the adversary does. Here, we consider the different but related problem of *authentication*, based on the observation that it is better to know what you do not know than it is to be wrong. We assume that an adversary may or may not be present; if it is not present, then the receiver should be able to decode the message at the maximum possible rate; if it is present, then the receiver should be able to *detect* the adversary, even if it cannot decode the message. Thus, when a message is decoded, the receiver can be certain that this message is authentic. Moreover, this must be accomplished without any pre-shared key.

Authentication capacity has been previously considered in [2], [3]. In [2], an inner bound is given when the adversary is assumed to know the code and the message in real-time. In our previous work [3], the authentication problem appeared in the context of establishing connectivity conditions in a joint compound-channel/AVC network. In the latter, the adversary is assumed to know the code but not the message. An alternative model was considered in [4], wherein the decoder receives either a noisy version of either the legitimate or the adversarial transmission, rather than a combination.

In the present paper, we adopt the model of [3], wherein the adversary knows the code but not the message, and exactly characterize the authentication capacity for any discrete-memoryless adversarial channel. Namely, we show that the authentication capacity is either zero or equal to the no-adversary capacity, depending on whether the channel is *overwritable*. Overwritability, a condition that we introduced in [3], is analogous to *symmetrizability*, which is a critical condition found in [5] for the classical AVC problem. Symmetrizability captures the condition wherein the adversary can inject a second, legitimate-appearing message. In contrast, overwritability captures the stronger condition wherein the adversary can completely replace the legitimate message with a counterfeit, without giving away its presence. Thus, authentication is impossible for an overwritable channel. We further prove that for any non-overwritable channel, authentication can be achieved without any loss in capacity compared to the no-adversary setting. Our achievability proof follows along the lines of [6], which established that the AVC capacity is either zero or equal to the random code capacity, since any positive rate code can be used to establish shared randomness between encoder and decoder. Similarly, we first show that for any non-overwritable channel, the authentication capacity is positive; this proof (Lemma 4) makes use of a technique similar to [5]. Subsequently, we show that a small amount of shared randomness is enough to achieve the no-adversary capacity, so any positive rate code can be used to establish shared randomness, and thereby achieve capacity.

II. Problem Description

Notation: Sequences of length n are denoted by bold symbols, such as $\mathbf{x}$. The type of a sequence $\mathbf{x} \in \mathcal{X}^n$ is denoted $P_{\mathbf{x}}(x) = \frac{1}{n}|\{i : x_i = x\}|$. Type classes will be written in terms of a random variable with distribution equal to the type. In particular, given a random variable X with alphabet $\mathcal{X}$ and distribution P_X, where $P_X(x)$ is an integer multiple of $1/n$

This material is based upon work supported by the National Science Foundation under grants CCF-1422358 and CNS-1526547, and also by the Army Research Office under contracts W911NF-17-2-0208 and W911NF-17-2-0183.

for all $x \in \mathcal{X}$, we write τ_X as the type class corresponding to P_X, i.e., $\tau_X = \{\mathbf{x} : P_{\mathbf{x}} = P_X\}$. Note that $P_{\mathbf{x}}$ is the type of the sequence $\mathbf{x}$, whereas P_X is simply the distribution of the single-letter random variable X. For any distribution such as P_X or conditional distribution such as $P_{Y|X}$, let P_X^n and $P_{Y|X}^n$ be the n-length memoryless extensions. The probability measure is denoted $\mathbb{P}(\cdot)$. All exponentials and logarithms have base 2. For an integer n, $[n] = \{1, \dots, n\}$.

Consider a discrete adversarial channel given by $W(y|x,s)$ with finite alphabets $\mathcal{X}, \mathcal{Y}, \mathcal{S}$ for input, output, and state (i.e., adversary signal) respectively. We denote a special symbol $s_0 \in \mathcal{S}$ as the *no-adversary state*. We define an (M, n) *authentication code* as an encoder/decoder pair

$$f : \{1, \dots, M\} \to \mathcal{X}^n \tag{1}$$

$$\phi : \mathcal{Y}^n \to \{0, 1, \dots, M\}. \tag{2}$$

A decoder output of 0 indicates an error declaration; i.e., that the decoder has detected the adversary. The goal is that if no adversary is present (i.e., the state sequence is $\mathbf{s} = \mathbf{s}_0 := (s_0, \dots, s_0)$), then the message should reliably decoded, but if an adversary is present (i.e., if $\mathbf{s} \neq \mathbf{s}_0$), then either the message is decoded correctly or the adversary is detected. Thus, given message i and state sequence $\mathbf{s}$, we define the probability of error for authentication code (f, ϕ) as

$$e(i, \mathbf{s}) = \begin{cases} W^n(\phi^{-1}(i)^c | f(i), \mathbf{s}_0), & \text{if } \mathbf{s} = \mathbf{s}_0 \\ W^n(\phi^{-1}(\{0, i\})^c | f(i), \mathbf{s}), & \text{if } \mathbf{s} \neq \mathbf{s}_0 \end{cases} \tag{3}$$

where $\phi^{-1}(\mathcal{A})^c$ is the set of $\mathbf{y} \in \mathcal{Y}^n$ such that $\phi(\mathbf{y}) \notin \mathcal{A}$. The average probability of error for state sequence $\mathbf{s}$ is

$$e(\mathbf{s}) = \frac{1}{M} \sum_{i=1}^{M} e(i, \mathbf{s}). \tag{4}$$

We say a rate R is *achievable* if there exists a sequence of $(2^{nR}, n)$ authentication codes with

$$\max_{\mathbf{s}} e(\mathbf{s}) \to 0 \text{ as } n \to \infty. \tag{5}$$

Note the implicit assumption that the adversary's choice of $\mathbf{s}$ may depend on the code (f, ϕ), but not the specific message selected at the encoder. The *authentication capacity* C_{auth} is the supremum of all achievable rates.

III. Main Results

The no-adversary capacity is given by

$$C = \max_{P_X} \; I(X; Y | S = s_0). \tag{6}$$

The random code capacity for an AVC [1] is given by

$$C_R = \max_{P_X} \min_{P_S} \; I(X; Y). \tag{7}$$

An AVC $W(y|x,s)$ is *symmetrizable* [5] if there exists a distribution $P_{S|X}$ such that

$$\sum_s P_{S|X}(s|x') W(y|x,s) = \sum_s P_{S|X}(s|x) W(y|x',s) \quad \text{for all } x, x', y. \tag{8}$$

An adversarial channel $W(y|x,s)$ with no-adversary state s_0 is *overwritable* [3] if there exists a distribution $P_{S|X'}$ such that

$$\sum_s P_{S|X'}(s|x') W(y|x,s) = W(y|x', s_0) \text{ for all } x, x', y. \tag{9}$$

That is, in an overwritable channel, the adversary can make it appear that it is not present, and that the input is whatever it chooses. The following proposition, proved in the appendix, gives simple relationships between these properties.

Proposition 1: The following properties are ordered from strongest to weakest (i.e., each implies the next):

1) Channel $W(y|x,s)$ has zero no-adversary capacity.
2) Channel $W(y|x,s)$ is overwritable.
3) Channel $W(y|x,s)$ has zero random code capacity.
4) Channel $W(y|x,s)$ is symmetrizable.

The following is our main result, giving the exact authentication capacity.

Theorem 2: If the channel is not overwritable, $C_{\text{auth}} = C$; if it is overwritable, then $C_{\text{auth}} = 0$.

IV. Examples

The following three example channels are each noiseless p-ary channels if no adversary is present (thus the no-adversary capacities are $C = \log p$), but have different characteristics in the context of the adversary.

Example 1: Let $\mathcal{X} = \mathcal{Y} = \{0, \dots, p-1\}$ and $\mathcal{S} = \{\mathsf{e}, 0, \dots, p-1\}$, where $s_0 = \mathsf{e}$. The output is a deterministic function of the input and the state, where $Y = X$ if $S = \mathsf{e}$, and $Y = S$ if $S \neq \mathsf{e}$. This channel is evidently overwritable, as the adversary may simply take $S = X'$. Thus, by Prop. 1, $C_R = 0$, and it also symmetrizable.

Example 2: Let $\mathcal{X} = \mathcal{Y} = \mathcal{S} = \{0, \dots, p-1\}$, where $s_0 = 0$ and $Y = X + S \bmod p$. This channel is not overwritable, but $C_R = 0$, so by Prop. 1 it is also symmetrizable.

Example 3: Let $\mathcal{X} = \mathcal{S} = \{0, \dots, p-1\}$, $\mathcal{Y} = \{0, \dots, 2(p-1)\}$, where $s_0 = 0$ and $Y = X + S$ with real addition. This channel is not overwritable, and has positive random code capacity $C_R = (1 - \frac{1}{p}) \log p - \frac{2}{p^2} \sum_{j=1}^{p-1} j \log j$, but it is symmetrizable.

V. Proofs

We prove Thm. 2 via four lemmas, stated and proved below. The structure of these lemmas is summarized as follows:

- Lemma 3 gives the converse for Thm. 2.
- Lemma 4 asserts the existence of positive rate codes for non-overwritable channels. Despite the seeming simplicity of proving the achievability of any positive rate, this lemma is the core piece of the achievability argument.
- Lemma 5 shows that the no-adversary capacity C is achievable with authentication if the encoder and decoder have access to a small amount of shared randomness, unknown to the adversary.
- Lemma 6 completes the achievability proof by using the low-rate code of Lemma 4 to establish shared randomness, and then the high-rate code of Lemma 5 to achieve C.

Lemma 3: For any adversarial channel, $C_{\text{auth}} \le C$. For an overwritable channel, $C_{\text{auth}} = 0$.

Proof: An authentication code must be a standard error-correcting channel code in the no-adversary case, so the classical converse bound gives $C_{\text{auth}} \le C$.

Consider an overwritable channel. Let $P_{S|X'}$ be the conditional distribution asserted by (9). Consider any sequence of $(2^{nR}, n)$ authentication codes. For any distribution $Q_{\mathbf{S}}$ on $\mathcal{S}^n$, we may bound $\max_{\mathbf{s}} e(\mathbf{s}) \ge \sum_{\mathbf{s}} Q_{\mathbf{S}}(\mathbf{s}) e(\mathbf{s})$. Specifically, let $M = 2^{nR}$ and $Q_{\mathbf{S}}(\mathbf{s}) = \frac{1}{M}\sum_{j=1}^{M} P^n_{S|X'}(\mathbf{s}|f(j))$. Thus

$$\max_{\mathbf{s}} e(\mathbf{s}) \ge \sum_{\mathbf{s}} \frac{1}{M}\sum_{j=1}^{M} P^n_{S|X'}(\mathbf{s}|f(j)) e(\mathbf{s}) \tag{10}$$

$$\ge \sum_{\mathbf{s}} \frac{1}{M}\sum_{j=1}^{M} P^n_{S|X'}(\mathbf{s}|f(j)) \frac{1}{M}\sum_{i=1}^{M} W^n(\phi^{-1}(\{0,i\})^c | f(i), \mathbf{s})$$

$$= \frac{1}{M^2}\sum_{i,j}\sum_{\mathbf{s}} P^n_{S|X'}(\mathbf{s}|f(j))\, W^n(\phi^{-1}(\{0,i\})^c | f(i), \mathbf{s})$$

$$= \frac{1}{M^2}\sum_{i,j} W^n(\phi^{-1}(\{0,i\})^c | f(j), \mathbf{s}_0) \tag{11}$$

$$\ge \frac{1}{M^2}\sum_{j=1}^{M}\sum_{i\neq j} W^n(\phi^{-1}(j) | f(j), \mathbf{s}_0) \tag{12}$$

$$= \frac{M-1}{M}(1 - e(\mathbf{s}_0)) \tag{13}$$

where (11) holds by the overwritability condition (9). Rearranging gives

$$(1 + \tfrac{M-1}{M}) \max_{\mathbf{s}} e(s) \ge \max_{\mathbf{s}} e(s) + \tfrac{M-1}{M} e(\mathbf{s}_0) \ge \tfrac{M-1}{M}$$

so $\max_{\mathbf{s}} e(s) \ge \frac{M-1}{2M-1}$. In particular, for any $R > 0$, the probability of error is bounded away from 0. Therefore, $C_{\text{auth}} = 0$. ∎

Lemma 4: For any non-overwritable channel, $C_{\text{auth}} > 0$.

Proof: Fix positive constant ϵ to be determined. Let P_X be any n-length type where $P_X(x) > 0$ for all $x \in \mathcal{X}$. We construct a $(2^{n\epsilon}, n)$ code with vanishing probability of error as follows.

Codebook: By Lemma 3 of [5], there exist codewords $\mathbf{x}_1, \ldots, \mathbf{x}_M$ of type P_X, where $M = 2^{n\epsilon}$ such that, for any $\mathbf{x}$, $\mathbf{s}$, and every joint type $P_{XX'S}$,

$$\frac{1}{M}|\{i : (\mathbf{x}_i, \mathbf{s}) \in \tau_{XS}\}| \le \exp\{-n\epsilon/2\} \quad \text{if } I(X;S) > \epsilon \tag{14}$$

$$\frac{1}{M}|\{i : (\mathbf{x}_i, \mathbf{x}_j, \mathbf{s}) \in \tau_{XX'S} \text{ for some } j \neq i\}| \le \exp\{-n\epsilon/2\} \quad \text{if } I(X;X'S) > 2\epsilon. \tag{15}$$

Encoding: Given any message $i \in [M]$, send $\mathbf{x}_i$.

Decoding: For any η, let

$$\mathscr{C}_\eta = \{P_{XSY} : D(P_{XSY} \| P_X \times P_S \times W) \le \eta\}. \tag{16}$$

Given output sequence $\mathbf{y}$, decode to message i if

1) the joint type $P_{\mathbf{x}_i, \mathbf{s}_0, \mathbf{y}} \in \mathscr{C}_{2\epsilon}$
2) For any other $j \neq i$ for which $P_{\mathbf{x}_j, \mathbf{s}', \mathbf{y}} \in \mathscr{C}_{2\epsilon}$ for some $\mathbf{s}'$, we have $I(XY; X') \le 5\epsilon$ where $XX'Y$ are dummy variables with $(\mathbf{x}_i, \mathbf{x}_j, \mathbf{y}) \in \tau_{XX'Y}$.

If there is no such message or more than one, declare an error.

Probability of error analysis: For any message i and state sequence $\mathbf{s}$, let $e_1(i, \mathbf{s})$ be the probability that message i does not satisfy the decoding requirement assuming message i is transmitted. Let $e_2(i, \mathbf{s})$ be the probability that some false message $j \neq i$ satisfies the decoding requirement assuming message i is transmitted. We need to show the following:

$$\frac{1}{M}\sum_{i=1}^{M} e_1(i, \mathbf{s}_0) \to 0 \tag{17}$$

$$\frac{1}{M}\sum_{i=1}^{M} e_2(i, \mathbf{s}) \to 0 \text{ for all } \mathbf{s}. \tag{18}$$

For each message $i \in [M]$, define the following events:

$$\mathcal{E}_{1i} = \{P_{\mathbf{x}_i, \mathbf{s}, \mathbf{Y}} \notin \mathscr{C}_{2\epsilon}\} \tag{19}$$

$$\mathcal{E}_{2i} = \{\exists j \neq i,\, \mathbf{s}' : (\mathbf{x}_i, \mathbf{x}_j, \mathbf{s}', \mathbf{Y}) \in \tau_{XX'S'Y} \text{ where } P_{X'S'Y} \in \mathscr{C}_{2\epsilon} \text{ and } I(XY; X') > 5\epsilon\}. \tag{20}$$

Note that message i satisfies the decoding requirement iff $\mathcal{E}_{1i}^c \cap \mathcal{E}_{2i}^c$, we may rewrite the two error probabilities by

$$e_1(i, \mathbf{s}) = \mathbb{P}\{\mathcal{E}_{1i} \cup \mathcal{E}_{2i} | \mathbf{X} = \mathbf{x}_i\} \tag{21}$$

$$e_2(i, \mathbf{s}) = \mathbb{P}\left\{\bigcup_{j\neq i} (\mathcal{E}_{1j}^c \cap \mathcal{E}_{2j}^c) \,\middle|\, \mathbf{X} = \mathbf{x}_i\right\}. \tag{22}$$

We define the following sets:

$$\mathcal{A}_1(\mathbf{s}) = \{i : (\mathbf{x}_i, \mathbf{s}) \in \tau_{XS} \text{ where } I(X;S) > \epsilon\} \tag{23}$$

$$\mathcal{A}_2(\mathbf{s}) = \{i : \exists j \neq i \text{ such that } (\mathbf{x}_i, \mathbf{x}_j, \mathbf{s}) \in \tau_{XX'S} \text{ where } I(X;X'S) > 2\epsilon\}. \tag{24}$$

For any $\mathbf{s}$, we may bound

$$\frac{1}{M}|\mathcal{A}_1(\mathbf{s})| = \frac{1}{M}\sum_{P_{XS}: I(X;S) > \epsilon} |\{i : (\mathbf{x}_i, \mathbf{s}) \in \tau_{XS}\}| \tag{25}$$

$$\le \sum_{P_{XS}: I(X;S) > \epsilon} \exp\{-n\epsilon/2\} \tag{26}$$

$$\le \exp\{-n\epsilon/3\} \tag{27}$$

where (26) follows from (14) and (27) holds for sufficiently large n because there are polynomially many types. By similar reasoning using (15),

$$\frac{1}{M}|\mathcal{A}_2(\mathbf{s})| \le \exp\{-n\epsilon/3\}. \tag{28}$$

For any $i \notin \mathcal{A}_1(\mathbf{s})$, we have

$$\mathbb{P}\{\mathcal{E}_{1i} | \mathbf{X} = \mathbf{x}_i\} = \sum_{\substack{P_{XSY} \notin \mathscr{C}_{2\epsilon}: \\ (\mathbf{x}_i, \mathbf{s}) \in \tau_{XS}}} \sum_{\mathbf{y}: (\mathbf{x}_i, \mathbf{s}, \mathbf{y}) \in \tau_{XSY}} W^n(\mathbf{y} | \mathbf{x}_i, \mathbf{s})$$

$$\le \sum_{\substack{P_{XSY} \notin \mathscr{C}_{2\epsilon}: \\ (\mathbf{x}_i, \mathbf{s}) \in \tau_{XS}}} \exp\{-nD(P_{XSY} \| P_{XS} \times W)\} \tag{29}$$

$$= \sum_{\substack{P_{XSY}\notin\mathscr{C}_{2\epsilon}:\\ (\mathbf{x}_i,\mathbf{s})\in\tau_{XS}}} \exp\{-n(D(P_{XSY}\|P_X\times P_S\times W)-I(X;S))\}$$

$$\leq \sum_{\substack{P_{XSY}\notin\mathscr{C}_{2\epsilon}:\\ (\mathbf{x}_i,\mathbf{s})\in\tau_{XS}}} \exp\{-n\epsilon\} \tag{30}$$

$$\leq \exp\{-n\epsilon/2\} \tag{31}$$

where (29) follows by basic facts about the method of types, (30) follows because $i \notin \mathcal{A}_1(\mathbf{s})$ and $P_{XSY} \notin \mathscr{C}_{2\epsilon}$, and (31) holds for sufficiently large n because there are only polynomially many types.

Following [5], for any random variables X, X', S, Y, $i \in [M]$, and $\mathbf{s}$, let

$$e_{XX'SY}(i,\mathbf{s}) = \sum_{\substack{y:(\mathbf{x}_i,\mathbf{x}_j,\mathbf{s},\mathbf{y})\in\tau_{XX'SY}\\ \text{for some } j\neq i}} W^n(\mathbf{y}|\mathbf{x}_i,\mathbf{s}). \tag{32}$$

By an identical argument as in [5],

$$e_{XX'SY}(i,\mathbf{s}) \leq \exp\{-n(I(Y;X'|XS)-3\epsilon)\}. \tag{33}$$

For $\mathbf{s}=\mathbf{s}_0$, and $i \in \mathcal{A}_1(\mathbf{s}_0)^c \cap \mathcal{A}_2(\mathbf{s}_0)^c$ we may now write

$$\mathbb{P}\{\mathcal{E}_{1i}^c \cap \mathcal{E}_{2i} | \mathbf{X}=\mathbf{x}_i\} \leq \sum_{\substack{P_{XX'SY}:P_{Xs_0Y}\in\mathscr{C}_{2\epsilon}\\ P_{X'S'Y}\in\mathscr{C}_{2\epsilon} \text{ for some } S'\\ I(XY;X')>5\epsilon}} e_{XX'SY}(i,\mathbf{s}_0) \tag{34}$$

For any term in this sum, we may apply (33) for the case $S=s_0$ to find

$$e_{XX'SY}(i,\mathbf{s}_0) \leq \exp\{-n(I(Y;X'|X)-3\epsilon)\} \tag{35}$$

$$\leq \exp\{-n(I(XY;X')-4\epsilon)\} \tag{36}$$

$$< \exp\{-n\epsilon\} \tag{37}$$

where (36) follows since $i \in \mathcal{A}_2(\mathbf{s}_0)^c$, so $I(X';X) \leq \epsilon$; and (37) follows since $I(XY;X') > 5\epsilon$. Thus, from (34), (37), and the fact that there are polynomially many types, for sufficiently large n,

$$\mathbb{P}\{\mathcal{E}_{1i}^c \cap \mathcal{E}_{2i} | \mathbf{X}=\mathbf{x}_i\} \leq \exp\{-n\epsilon/2\}. \tag{38}$$

Combining (27), (28), (31), and (38) proves (17).

It remains to prove (18). For any message i and state sequence $\mathbf{s}$, we may write

$$e_2(i,\mathbf{s}) \leq \mathbb{P}\{\mathcal{E}_{1i}|\mathbf{X}=\mathbf{x}_i\} + \mathbb{P}\left\{\mathcal{E}_{1i}^c \cap \bigcup_{j\neq i}(\mathcal{E}_{1j}^c \cap \mathcal{E}_{2j}^c) \,\middle|\, \mathbf{X}=\mathbf{x}_i\right\}. \tag{39}$$

The first term in (39) is bounded by (31). For the second term, assuming $\mathcal{E}_{1i}^c$, in order for $\mathcal{E}_{2j}^c$ to occur, it must be the case that $(\mathbf{x}_i,\mathbf{x}_j,\mathbf{s},\mathbf{y}) \in \tau_{XX'SY}$ where $I(X'Y;X) \leq 5\epsilon$. Thus we may upper bound the second term in (39) by

$$\sum_{\substack{P_{XX'SY}:P_{X's_0Y}\in\mathscr{C}_{2\epsilon}\\ P_{XSY}\in\mathscr{C}_{2\epsilon}\\ I(X'Y;X)\leq 5\epsilon}} e_{XX'SY}(i,\mathbf{s}) \tag{40}$$

Assume $i \in \mathcal{A}_1(\mathbf{s})^c \cap \mathcal{A}_2(\mathbf{s})^c$, and consider any $P_{XX'SY}$ in the above sum. We claim that $I(Y;X'|XS) \geq 4\epsilon$ for ϵ sufficiently small. Suppose not. Then, putting together several of the above, the joint distribution must satisfy

$$\begin{aligned} I(Y;X'|XS) &< 4\epsilon, & P_{XSY} &\in \mathscr{C}_{2\epsilon},\\ I(X;X'S) &\leq \epsilon, & P_{X's_0Y} &\in \mathscr{C}_{2\epsilon}\\ I(X'Y;X) &\leq 5\epsilon. \end{aligned} \tag{41}$$

In the limit as $\epsilon \to 0$, we would have

$$P_{XX'SY}(x,x',s,y) = P_X(x)P_X(x')P_{S|X'}(s|x')W(y|x,s),$$
$$P_{X'Y}(x',y) = P_X(x')W(y|x',s_0).$$

These conditions are precisely those that define overwritable. Since by assumption the channel is not overwritable, by continuity, for sufficiently small ϵ, (41) cannot hold simultaneously. This proves that $I(Y;X'|XS) \geq 4\epsilon$, so by (33), $e_{XX'SY}(i,\mathbf{s}) \leq \exp\{-n\epsilon\}$. Therefore, as there are polynomially many types, for sufficiently large n the second term in (39) is at most $\exp\{-n\epsilon/2\}$. This proves (18). ∎

Lemma 5: Fix a type P_X, rate $R < I(X;Y|S=s_0)$, and constant ϵ. Also let δ be a positive constant such that $R+\delta < I(X;Y|S=s_0)$. Let K be an integer such that

$$K > \frac{2n}{\epsilon}(R+\log|\mathcal{S}|). \tag{42}$$

Let $M = 2^{nR}$. There exist codewords $\mathbf{x}_{ik}$ for $i \in [M]$ and $k \in [K]$, all of type P_X, such that the following holds. For each $k \in [K]$, define a decoder $\phi_k(\mathbf{y})$ as follows:

- if i is the unique integer such that $(\mathbf{x}_i,\mathbf{y}) \in \tau_{XY}$ for some P_{XY} satisfying $I(X;Y) \geq R+\delta$, set $\phi_k(\mathbf{y}) = i$
- if there is no such integer or more than one, set $\phi_k(\mathbf{y}) = 0$ (i.e. declare an error).

The following hold:

$$\frac{1}{K}\sum_{k=1}^{K} W^n(\phi_k^{-1}(i)^c|\mathbf{x}_{ik},\mathbf{s}_0) \leq \epsilon \text{ for all } i \in [M] \tag{43}$$

$$\frac{1}{K}\sum_{k=1}^{K} W^n(\phi_k^{-1}(\{0,i\})^c|\mathbf{x}_{ik},\mathbf{s}) \leq \epsilon \quad \text{for all } i \in [M] \text{ and all } \mathbf{s}\neq\mathbf{s}_0. \tag{44}$$

Proof: We first prove that any rate $R < I(X;Y|S=s_0)$ is achievable by an authentication code in which the encoder and decoder have access to arbitrary amounts of shared randomness, unknown to the adversary. We then use a random code reduction, identical to that used for the AVC (in particular, Lemma 12.8 of [7]) to show that the same rate can be achieved with much less shared randomness.

We define a randomized code (F,Φ), independent of the adversary, as follows. Let $\delta > 0$ be small enough so that $R+\delta < I(X;Y|S=s_0)$. Let $\mathbf{X}_1,\ldots,\mathbf{X}_M$ be drawn independently and uniformly from τ_X. This random codebook constitutes the shared randomness between encoder and decoder. Set $F(i) = \mathbf{X}_i$ for all $i \in [M]$ and let $\Phi(\mathbf{y}) = i$ if i is the unique integer for which $(\mathbf{X}_i,\mathbf{y}) \in \tau_{XY}$ where

$I(X;Y) \geq R+\delta$; and $\Phi(\mathbf{y}) = 0$ if there is no such integer or more than one.

Fix some message i and state sequence $\mathbf{s}$. Let $\mathbf{Y} \sim W^n(\mathbf{y}|\mathbf{X}_i, \mathbf{s})$. We will prove the following:

$$\mathbb{P}\{(\mathbf{X}_i, \mathbf{Y}) \in \tau_{XY} \text{ where } I(X;Y) < R+\delta\} \to 0 \text{ for } \mathbf{s} = \mathbf{s}_0 \tag{45}$$

$$\mathbb{P}\{(\mathbf{X}_j, \mathbf{Y}) \in \tau_{XY} \text{ where } I(X;Y) \geq R+\delta \text{ for some } j \neq i\} \to 0 \text{ for all } \mathbf{s}. \tag{46}$$

Letting $\mathbf{s} = \mathbf{s}_0$, we may write the probability in (45) as

$$\sum_{P_{XY}: I(X;Y)<R+\delta} \; \sum_{(\mathbf{x},\mathbf{y})\in\tau_{XY}} |\tau_X|^{-1} W^n(\mathbf{y}|\mathbf{x}, \mathbf{s}_0) \tag{47}$$

$$\leq \sum_{P_{XY}: I(X;Y)<R+\delta} \; \sum_{\mathbf{x}\in\tau_X} |\tau_X|^{-1} \cdot \exp\{-nD(P_{XY}\|P_X \times W_{Y|X,s_0})\} \tag{48}$$

$$= \sum_{P_{XY}: I(X;Y)<R+\delta} \exp\{-nD(P_{XY}\|P_X \times W_{Y|X,s_0})\} \tag{49}$$

Note that if $P_{XY} = P_X \times W_{Y|X,s_0}$, then $I(X;Y) = I(X;Y|S=s_0) > R+\delta$. Since the relative entropy and mutual information are continuous functions of P_{XY}, there exists $\epsilon' > 0$ such that, if $I(X;Y) < R+\delta$, then

$$D(P_{XY}\|P_X \times W_{Y|X,s_0}) \geq \epsilon'. \tag{50}$$

Thus, since there are polynomially many types, for sufficiently large n, the probability in (45) is at most $\exp\{-n\epsilon'/2\}$. This proves (45).

For any $\mathbf{s}$, we may write the probability in (46) as

$$\sum_{P_{XX'SY}: I(X';Y)\geq R+\delta} \tilde{e}_{XX'SY} \tag{51}$$

where the sum only includes joint types $P_{XX'SY}$ where $P_{X'} = P_X$ (i.e., the type chosen at the outset), and where

$$\tilde{e}_{XX'SY} = \sum_{\mathbf{x}_1,\ldots,\mathbf{x}_M} |\tau_X|^{-M} \sum_{\substack{\mathbf{y}:(\mathbf{x}_i,\mathbf{x}_j,\mathbf{s},\mathbf{y})\in\tau_{XX'SY} \\ \text{for some } j\neq i}} W^n(\mathbf{y}|\mathbf{x}_i, \mathbf{s})$$

$$\leq \sum_{\mathbf{x}_1,\ldots,\mathbf{x}_M} |\tau_X|^{-M} \sum_{j\neq i} \sum_{\mathbf{y}:(\mathbf{x}_i,\mathbf{x}_j,\mathbf{s},\mathbf{y})\in\tau_{XX'SY}} W^n(\mathbf{y}|\mathbf{x}_i, \mathbf{s}) \tag{52}$$

$$= \sum_{\substack{\mathbf{x},\mathbf{x}',\mathbf{y}: \\ (\mathbf{x},\mathbf{x}',\mathbf{s},\mathbf{y})\in\tau_{XX'SY}}} (M-1)|\tau_X|^{-2} \; W^n(\mathbf{y}|\mathbf{x}, \mathbf{s}) \tag{53}$$

$$\leq \exp\{n(R - 2H(X) + H(XX'Y|S) - H(Y|XS) + \epsilon)\} \tag{54}$$

$$= \exp\{n(R - I(X;S) - I(X';XSY) + \epsilon)\} \tag{55}$$

$$\leq \exp\{n(R - I(X';Y) + \epsilon)\} \tag{56}$$

$$\leq \exp\{n(-\delta + \epsilon)\} \tag{57}$$

where in (55) we have used the fact that $H(X) = H(X')$. Since there are polynomially many types, (51) is vanishing in n as long as $\delta > \epsilon$. This proves (46).

We have shown that the randomized code (F, Φ) has vanishing probability of error. Applying Lemma 12.8 of [7] to the randomized code (F, Φ) completes the proof of the lemma. ■

Lemma 6: For a non-overwritable channel, $C_{\text{auth}} \geq C$.

Proof: Fix any $R < C$ and any $\epsilon > 0$. Let $M = \exp\{nR\}$ and $K = n^2$. For sufficiently large n there exists an n-length type P_X where $R < I(X;Y|S=s_0)$. Also, for sufficiently large n, (42) is satisfied by the choice of $K = n^2$.

Codebooks: Apply Lemma 5 to assert the existence of codewords $\mathbf{x}_{ik}$ for $i \in [M]$ and $k \in [K]$, as well as the associated decoders ϕ_k.

Encoding: Let the message consist of the pair (k, i) where $k \in [K]$, and $i \in [M]$. Using any code of positive-rate (whose existence is asserted by Lemma 4), first encode k. Then transmit $\mathbf{x}_{ij}$.

Decoding: Let $\hat{k}$ be the estimate of k bound by the decoder from Lemma 4. If $\hat{k} = 0$ (i.e., the code declares an error), then declare an error for the overall code. Otherwise, let $\mathbf{y}$ be the n-length sequence associated with the codeword $\mathbf{x}_{ij}$, and find $\hat{i} = \phi_{\hat{k}}(\mathbf{y})$. If $\hat{i} = 0$ (i.e., the second code declares an error), we declare an error. Otherwise the decoded message is $(\hat{k}, \hat{i})$.

Probability of error analysis: By Lemma 4, with high probability, $\hat{k} = k$ if no adversary is present, or an error is declared. Assuming k is decoded correctly, the probability of error for both the $\mathbf{s} = \mathbf{s}_0$ and $\mathbf{s} \neq \mathbf{s}_0$ cases are given by the quantities in (43)–(44). Thus the probability of error is bounded by ϵ. ■

Appendix: Proof of Proposition 1

An AVC has zero no-adversary capacity if and only if there exists a distribution P_Y such that $W(y|x, s_0) = P_Y(y)$ for all x, y. Thus, if 1) holds, we may take $P_{S|X'} = 1(s = s_0)$, so $\sum_s P_{S|X'}(s|x')W(y|x,s) = W(y|x,s_0) = W(y|x',s_0)$. Thus, 1) implies 2). To prove that 2) implies 3), fix any $\tilde{x}$, and let $P_S(s) = P_{S|X'}(s|\tilde{x})$. Thus $\sum_s P_S(s)W(y|x,s) = W(y|\tilde{x},s_0)$. As the right-hand side does not depend on x, Y is independent of X, and hence $I(X;Y) = 0$. Thus the random code capacity is zero. That 3) implies 4) is shown in [5], and follows because (8) holds with the choice $P_{S|X}(s|x) = P_S(s)$, where P_S is a distribution giving $I(X;Y) = 0$.

References

[1] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacities of certain channel classes under random coding," *The Annals of Mathematical Statistics*, vol. 31, no. 3, pp. 558–567, 1960.

[2] E. Graves, P. Yu, and P. Spasojevic, "Keyless authentication in the presence of a simultaneously transmitting adversary," in *2016 IEEE Information Theory Workshop (ITW)*, Sept 2016, pp. 201–205.

[3] O. Kosut and J. Kliewer, "Network equivalence for a joint compound-arbitrarily-varying network model," in *2016 IEEE Information Theory Workshop (ITW)*, Sept 2016, pp. 141–145.

[4] J. Perazzone, E. Graves, P. Yu, and R. Blum, "Inner bound for the capacity region of noisy channels with an authentication requirement," [Online] arXiv:1801.03920, Jan. 2018.

[5] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel revisited: positivity, constraints," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 181–193, 1988.

[6] R. Ahlswede, "Elemination of correlation in random codes for arbitrarily varing channels," *Probability Theory and Related Fields*, vol. 44, no. 2, pp. 159–175, 1978.

[7] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.