Alternative security models for pseudorandom functions

Kirill Tsaregorodtsev Researcher at Cryptography laboratory, JSRPC ``Kryptonite'', Moscow, Russia

Криптонит

CTCrypt'2023

1. Introduction

- 2. Session key secrecy
- 3. Explicit authentication

4. User privacy

5. Conclusion

Introduction

Session key secrecy

Explicit authentication

User privacy

Conclusion

IK The origin of the problem

• Analysis of 5G-AKA protocol.

- Analysis of 5G-AKA protocol.
- We want: session key secrecy, explicit authentication, user privacy.

- Analysis of 5G-AKA protocol.
- We want: session key secrecy, explicit authentication, user privacy.
- These properties give rise to the different security models for the underlying pseudorandom functions (PRF).

K 5G-AKA in a nutshell

• Key agreement protocol based on a pre-shared secret keys.

IK 5G-AKA in a nutshell

- Key agreement protocol based on a pre-shared secret keys.
- Main part of the protocol: three messages.

₩ 5G-AKA in a nutshell

- Key agreement protocol based on a pre-shared secret keys.
- Main part of the protocol: three messages.



Ⅲ 5G-AKA: focusing on PRFs



• identify "correct" security properties needed for the reduction of 5G-AKA protocol;

- identify "correct" security properties needed for the reduction of 5G-AKA protocol;
- propose security models for PRF that formalizes these properties;

- identify "correct" security properties needed for the reduction of 5G-AKA protocol;
- propose security models for PRF that formalizes these properties;
- analyze obtained models; show that they can be reduced to the standard security model for PRF.

 $\mathcal{F} = \{\mathcal{F}_k \in Funs(Dom, Range) \mid k \in Keys\}$

$$\mathcal{F} = \{\mathcal{F}_k \in Funs(Dom, Range) \mid k \in Keys\}$$

Examples: block cipher "Magma"

$$\mathcal{F}_k(m) = E(k,m), Keys = \{0,1\}^{256}, Dom = Range = \{0,1\}^{64}, M = \{0,1\}^{64}$$

$$\mathcal{F} = \{\mathcal{F}_k \in Funs(Dom, Range) \mid k \in Keys\}$$

Examples: block cipher "Magma"

$$\mathcal{F}_k(m) = E(k,m), Keys = \{0,1\}^{256}, Dom = Range = \{0,1\}^{64}$$

MAC function MAC(k, \cdot) (in that case $Dom = \{0, 1\}^*$).

IK PRF model

The advantage of the adversary \mathcal{A} in the PRF model for the function family \mathcal{F} is the following quantity:

$$\mathrm{Adv}_{\mathcal{F}}^{\mathrm{PRF}}(\mathcal{A}) = \mathbb{P}\Big[\mathrm{Exp}_{\mathcal{F}}^{\mathrm{PRF}\text{-}1}(\mathcal{A}) \to 1\Big] - \mathbb{P}\Big[\mathrm{Exp}_{\mathcal{F}}^{\mathrm{PRF}\text{-}0}(\mathcal{A}) \to 1\Big].$$

K PRF model

The advantage of the adversary \mathcal{A} in the PRF model for the function family \mathcal{F} is the following quantity:

$$\mathrm{Adv}_{\mathcal{F}}^{\mathrm{PRF}}(\mathcal{A}) = \mathbb{P}\Big[\mathrm{Exp}_{\mathcal{F}}^{\mathrm{PRF}\text{-}1}(\mathcal{A}) \to 1\Big] - \mathbb{P}\Big[\mathrm{Exp}_{\mathcal{F}}^{\mathrm{PRF}\text{-}0}(\mathcal{A}) \to 1\Big].$$

 $\frac{\operatorname{Exp}_{\mathcal{F}}^{\operatorname{PRF}-1}(\mathcal{A})}{k \leftarrow^{\$} \operatorname{Keys}}$ $b' \leftarrow^{\$} \mathcal{A}^{\mathcal{O}_{\operatorname{prf}}}$ $\operatorname{return} b'$ $\mathcal{O}_{\operatorname{prf}}(m)$

return $\mathcal{F}_k(m)$

K PRF model

The advantage of the adversary \mathcal{A} in the PRF model for the function family \mathcal{F} is the following quantity:

$$\mathrm{Adv}_{\mathcal{F}}^{\mathrm{PRF}}(\mathcal{A}) = \mathbb{P}\Big[\mathrm{Exp}_{\mathcal{F}}^{\mathrm{PRF}\text{-}1}(\mathcal{A}) \to 1\Big] - \mathbb{P}\Big[\mathrm{Exp}_{\mathcal{F}}^{\mathrm{PRF}\text{-}0}(\mathcal{A}) \to 1\Big].$$

$\mathrm{Exp}_{\mathcal{F}}^{\mathrm{PRF}\text{-}1}(\mathcal{A})$	$\operatorname{Exp}_{\operatorname{\mathscr{F}}}^{\operatorname{PRF}\operatorname{-0}}(\operatorname{\mathscr{A}})$	$\mathcal{O}_{\mathrm{prf}}(m)$
$k \leftarrow^{\$} Keys$	Asked $\leftarrow []$	if $Asked[m] = \bot$
$b' \leftarrow^{\$} \mathcal{A}^{\mathcal{O}_{\mathrm{prf}}}$	$b' \leftarrow^{\$} \mathcal{A}^{\mathcal{O}_{\mathrm{prf}}}$	$Asked[m] \leftarrow^{\$} Range$
return b'	$\mathbf{return} \ b'$	fi
$\mathcal{O}_{\mathrm{prf}}(m)$		return Asked[m]
return $\mathcal{F}_k(m)$		

$\mathrm{Adv}_{\mathcal{F}}^{\mathrm{PRF}}(t,q,\ell,\mu)$

maximal advantage $\operatorname{Adv}_{\mathcal{F}}^{\operatorname{PRF}}(\mathcal{A})$, where the maximum is taken over the adversaries \mathcal{A} with

 $\mathrm{Adv}^{\mathrm{PRF}}_{\mathcal{F}}(t,q,\ell,\mu)$

maximal advantage $\operatorname{Adv}_{\mathcal{F}}^{\operatorname{PRF}}(\mathcal{A})$, where the maximum is taken over the adversaries \mathcal{A} with

- time complexity is at most *t*,
- \cdot the number of queries to $\mathcal{O}_{\mathrm{prf}}$ does not exceed q,
- · total length of the queries $\sum |m|$ does not exceed ℓ ,
- maximal query length $\max |m|$ does not exceed μ .

 $\mathcal{F}_k(x) = \operatorname{Hash}\left(k || \operatorname{Hash}\left(k || x\right)\right)$

Table 1: Calculation of values depending on the pre-shared secret k

Value	S3G function	Computation rule	Indices
σ_1	f_1	$\mathcal{F}_k(SQN \parallel RAND \parallel Const_1)$	[1: <i>tlen</i>]
σ_2	f_1^*	$\mathcal{F}_k(SQN_{UE} RAND Const_1)$	[257: 256 + tlen]
RES	f_2		[1: reslen]
AK	f_5	$\mathcal{F}_k(RAND \parallel Const_2)$	[257: 256 + 48]
AK^*	f_5^*		[305: 304 + 48]
СК	f_3		[1: klen]
		$\mathcal{F}_k(RAND \parallel Const_3)$	
IK	f_4		[257: 256 + klen]

• σ_1 , σ_2 (part of the *AUTN*, *AUTS* resp.): integrity of the transmitted messages within the session; explicit authentication Home Network and User resp.

- σ_1 , σ_2 (part of the *AUTN*, *AUTS* resp.): integrity of the transmitted messages within the session; explicit authentication Home Network and User resp.
- *RES*: explicit User authentication, confirmation of successful completion on the User's side.

- σ_1 , σ_2 (part of the *AUTN*, *AUTS* resp.): integrity of the transmitted messages within the session; explicit authentication Home Network and User resp.
- *RES*: explicit User authentication, confirmation of successful completion on the User's side.
- *AK*, *AK** (used in *AUTN*, *AUTS* resp.): pseudorandom sequence masking the connection counters *SQN*.

- σ_1 , σ_2 (part of the *AUTN*, *AUTS* resp.): integrity of the transmitted messages within the session; explicit authentication Home Network and User resp.
- *RES*: explicit User authentication, confirmation of successful completion on the User's side.
- *AK*, *AK** (used in *AUTN*, *AUTS* resp.): pseudorandom sequence masking the connection counters *SQN*.
- CK, IK: session key derivation k_{session}.



Introduction

Session key secrecy

Explicit authentication

User privacy

Conclusion

- High-level goal: obtaining information about the session key.
- **Goal (in model):** distinguish between a **segment** of a pseudorandom function output and a random string (in the presence of additional information).

- High-level goal: obtaining information about the session key.
- Goal (in model): distinguish between a segment of a pseudorandom function output and a random string (in the presence of additional information).
- **High-level capabilities:** compromise session keys in sessions other than the one being attacked, as well as receiving the values of σ_1 , σ_2 , *RES* (transmitted in plaintext) or partial information about the values of *AK*, *AK*^{*}.
- Capabilities (in model): learning output segments of a pseudorandom function.

IK PRF⁺ model: pseudocode

```
\operatorname{Exp}_{\varphi}^{\operatorname{PRF}^{+}-b}(\mathcal{A})
                                                                               \mathcal{O}_{\text{test}}^b(m, idx_1, idx_2)
k \leftarrow ^{\$} Kevs
                                                                                if (Asked[m] \cap [idx_1 : idx_2] \neq \emptyset)
Asked \leftarrow []
                                                                                    return \perp
h' \leftarrow {}^{\$} \mathcal{A}^{\mathcal{O}_{\mathrm{prf}},\mathcal{O}^{b}_{\mathrm{test}}}
                                                                               fi
                                                                               \mathbf{if}(b=0)
return b'
                                                                                   val \leftarrow {}^{\$} {0,1}^{idx_2 - idx_1 + 1}
\mathcal{O}_{\mathrm{prf}}(m, idx_1, idx_2)
                                                                                else
if (Asked[m] \cap [idx_1 : idx_2] \neq \emptyset)
                                                                                    val \leftarrow \mathcal{F}_{k}(m)[idx_{1}:idx_{2}]
    return \perp
                                                                               fi
fi
                                                                               Asked[m] \leftarrow Asked[m] \cup [idx_1 : idx_2]
Asked[m] \leftarrow Asked[m] \cup [idx_1 : idx_2]
                                                                                return val
return \mathcal{F}_k(m)[idx_1:idx_2]
```

$\mathrm{Adv}_{\mathcal{F}}^{\mathrm{PRF}^+}(t,q_{prf},q_{test})$ maximal value among $\mathrm{Adv}_{\mathcal{F}}^{\mathrm{PRF}^+}(\mathcal{A})$, where:

$\mathrm{Adv}_{\mathcal{F}}^{\mathrm{PRF}^+}(t,q_{prf},q_{test})$

maximal value among $\operatorname{Adv}_{\mathcal{F}}^{\operatorname{PRF}^+}(\mathcal{A})$, where:

- \cdot \mathcal{A} 's time complexity does not exceed t,
- \cdot $\, \mathcal{A} \,$ makes no more than q_{prf} queries to the $\mathcal{O}_{\mathrm{prf}}$,
- q_{test} queries to $\mathcal{O}^b_{ ext{test}}$ oracles.

The following inequality holds:

$$\operatorname{Adv}_{\mathcal{F}}^{\operatorname{PRF}^+}(t, q_{prf}, q_{test}) \leq 2 \cdot \operatorname{Adv}_{\mathcal{F}}^{\operatorname{PRF}}(t + q_{prf} + q_{test}, q_{prf} + q_{test}).$$

• "true" segments does not help much...

- "true" segments does not help much...
- because they are indistinguishable from random ones,

- "true" segments does not help much...
- because they are indistinguishable from random ones,
- \cdot hence, $\mathcal{O}_{\mathrm{prf}}$ can be excluded (i.e., modelled with a random string generator).

- "true" segments does not help much...
- because they are indistinguishable from random ones,
- $\cdot\,$ hence, $\mathcal{O}_{\rm prf}$ can be excluded (i.e., modelled with a random string generator).
- PRF⁺ model can be naturally generalized to the case of $D \in \mathbb{N}$ parties; by hybrid argument this case can be reduced to the case D = 1.

Introduction

Session key secrecy

Explicit authentication

User privacy

Conclusion

- High-level goal: explicit participant authentication.
- **Goal (in model):** forge the segment of a pseudorandom function output (in the presence of additional information).

- High-level goal: explicit participant authentication.
- Goal (in model): forge the segment of a pseudorandom function output (in the presence of additional information).
- **High-level capabilities:** compromise session keys in sessions other than the one being attacked, as well as receiving the values of σ_1 , σ_2 , *RES* (transmitted in plaintext) or partial information about the values of *AK*, *AK*^{*}.
- Capabilities (in model): learning output segments of a pseudorandom function.

IK UF-PRF model: pseudocode

$\operatorname{Exp}_{\operatorname{\mathscr{F}}}^{\operatorname{UF-PRF}}(\operatorname{\mathscr{A}})$	$\mathcal{O}_{\mathrm{vfy}}(m,\tau,i)$
$\overline{k \leftarrow^{\$} Keys}$	$val \leftarrow \mathcal{F}_k(m)[i: i + tlen - 1]$
Asked $\leftarrow []$	$res \leftarrow (\tau = val)$
$win \leftarrow false$	$\mathbf{if} \left(Asked[m] \cap [i: \ i+tlen-1] = \varnothing \right)$
$\mathcal{A}^{\mathcal{O}_{\mathrm{prf}},\mathcal{O}_{\mathrm{vfy}}}$	$win \leftarrow win \lor res$
return win	fi
$\mathcal{O}_{\mathrm{prf}}(m,idx_1,idx_2)$	return res
$Asked[m] \leftarrow Asked[m] \cup [idx_1: idx_2]$	
return $\mathcal{F}_k(m)[idx_1:idx_2]$	

$\mathrm{Adv}_{\mathcal{F}}^{\mathrm{UF}\mathrm{-}\mathrm{PRF}}(t,q_{prf},q_{vfy},tlen)$ maximal value among $\mathrm{Adv}_{\mathcal{F}}^{\mathrm{UF}\mathrm{-}\mathrm{PRF}}(\mathcal{A})$, where

$\mathrm{Adv}_{\mathcal{F}}^{\mathrm{UF} ext{-}\mathrm{PRF}}(t,q_{prf},q_{vfy},tlen)$ maximal value among $\mathrm{Adv}_{\mathcal{F}}^{\mathrm{UF} ext{-}\mathrm{PRF}}(\mathcal{A})$, where

- \cdot \mathcal{A} 's time complexity does not exceed t,
- the length of the segment to be predicted is *tlen*,
- \cdot $\, \mathcal{A} \,$ makes no more than q_{prf} queries to the $\mathcal{O}_{\mathrm{prf}}$,
- $\cdot \; q_{vfy}$ queries to $\mathcal{O}_{\mathrm{vfy}}$ oracles.

The following inequality holds:

$$\operatorname{Adv}_{\mathcal{F}}^{\operatorname{UF-PRF}}(t, q_{prf}, q_{vfy}, tlen) \leq \operatorname{Adv}_{\mathcal{F}}^{\operatorname{PRF}^+}(t + q_{prf} + q_{vfy}, q_{prf}, q_{vfy}) + \frac{q_{vfy}}{2tlen}.$$

• It is hard to distinguish segments from random ones...

¹Bellare, Goldreich, and Mityagin, The Power of Verification Queries in Message Authentication and Authenticated Encryption.

- It is hard to distinguish segments from random ones...
- hence, it is even harder to predict it completely¹...

¹Bellare, Goldreich, and Mityagin, The Power of Verification Queries in Message Authentication and Authenticated Encryption.

- It is hard to distinguish segments from random ones...
- hence, it is even harder to predict it completely¹...
- BUT: there is a chance to guess correctly $(\frac{q_{vfy}}{2^{tlen}}$ term).

¹Bellare, Goldreich, and Mityagin, The Power of Verification Queries in Message Authentication and Authenticated Encryption.

- It is hard to distinguish segments from random ones...
- hence, it is even harder to predict it completely¹...
- BUT: there is a chance to guess correctly $(\frac{q_{vfy}}{ztlen}$ term).
- UF-PRF model can be naturally generalized to the case of $D \in \mathbb{N}$ parties; by hybrid argument this model can be reduced to the case D = 1.

¹Bellare, Goldreich, and Mityagin, The Power of Verification Queries in Message Authentication and Authenticated Encryption.

Introduction

Session key secrecy

Explicit authentication

User privacy

Conclusion

- **High-level goal:** indistinguishable behaviour of users (cannot deduce which user is answering to the queries).
- **Goal (in model):** determine whether the adversary interacts with the "left" or "right" oracle (see also², LOR-DCPA model).

²Bellare, Kohno, and Namprempre, "Breaking and provably repairing the SSH authenticated encryption scheme: A case study of the Encode-then-Encrypt-and-MAC paradigm."

- **High-level goal:** indistinguishable behaviour of users (cannot deduce which user is answering to the queries).
- Goal (in model): determine whether the adversary interacts with the "left" or "right" oracle (see also², LOR-DCPA model).
- **High-level capabilities:** compromise session keys in sessions other than the one being attacked, as well as receiving the values of σ_1 , σ_2 , *RES* (transmitted in plaintext) or partial information about the values of *AK*, *AK*^{*}.
- Capabilities (in model): learning output segments of a pseudorandom function.

²Bellare, Kohno, and Namprempre, "Breaking and provably repairing the SSH authenticated encryption scheme: A case study of the Encode-then-Encrypt-and-MAC paradigm."

• The adversary has to determine on which of the keys ("left" k_{i_0} or "right" k_{i_1}) and which message ("left" m_0 or "right" m_1) is processed by the oracle \mathcal{O}_{lor}^b .

- The adversary has to determine on which of the keys ("left" k_{i_0} or "right" k_{i_1}) and which message ("left" m_0 or "right" m_1) is processed by the oracle \mathcal{O}_{lor}^b .
- To exclude the possibility of trivial attacks the adversary is not allowed to repeat messages for each fixed key k_i .

- The adversary has to determine on which of the keys ("left" k_{i_0} or "right" k_{i_1}) and which message ("left" m_0 or "right" m_1) is processed by the oracle \mathcal{O}_{lor}^b .
- To exclude the possibility of trivial attacks the adversary is not allowed to repeat messages for each fixed key k_i .
- In 5G-AKA message uniqueness is implemented by adding a counter *SQN* (number of connections) to the messages, as well as the randomness *RAND*.

$\operatorname{Exp}_{\operatorname{\mathscr{F}}}^{\operatorname{LOR-PRF}-b}(\operatorname{\mathscr{A}})$	$\mathcal{O}^b_{\mathrm{lor}}(m_0,i_0,m_1,i_1)$
for $i \in \{1, \dots, d\}$	if $(m_0 \in Msg[i_0]) \lor (m_1 \in Msg[i_1])$
$k_i \leftarrow^{\$} Keys$	$\mathbf{return} \perp$
$\mathbf{end}\mathbf{for}$	fi
$Msg \leftarrow []$	$Msg[i_0] \leftarrow Msg[i_0] \cup \{m_0\}$
$b' \leftarrow^{\$} \mathcal{A}^{\mathcal{O}^b_{\mathrm{lor}}}$	$Msg[i_1] \leftarrow Msg[i_1] \cup \{m_1\}$
${f return}\;b'$	$\mathbf{return} \; \mathcal{F}_{k_{i_b}}(m_b)$

$\operatorname{Adv}_{\mathcal{F}}^{\operatorname{LOR-PRF}}(t,Q;d)$

maximal value among $\operatorname{Adv}_{\mathcal{F}}^{\operatorname{LOR-PRF}}(\mathcal{A})$ in LOR-PRF Experiment with d users, where

$\mathrm{Adv}^{\mathrm{LOR}\text{-}\mathrm{PRF}}_{\mathcal{F}}(t,Q;d)$

maximal value among $\operatorname{Adv}_{\mathscr{F}}^{\operatorname{LOR-PRF}}(\mathscr{A})$ in LOR-PRF Experiment with d users, where

- \cdot \mathcal{A} 's time complexity does not exceed t,
- number of queries to \mathcal{O}^b_{lor} oracle on the key k_i (either as "left", or as "right", i.e., queries of the form (\cdot, i, \cdot, \cdot) or (\cdot, \cdot, \cdot, i) does not exceed Q[i].

$$\operatorname{Adv}_{\mathcal{F}}^{\operatorname{LOR-PRF}}(t,Q;d) \leq 2d \cdot \operatorname{Adv}_{\mathcal{F}}^{\operatorname{PRF}}(t+d+\sum_{i}Q[i],\max_{i}Q[i]).$$

 \cdot A series of hybrids $\mathcal{B}_{b_0}^{b_1,...,b_d}(\mathcal{A})$,

- A series of hybrids $\mathcal{B}_{b_0}^{b_1,\dots,b_d}(\mathcal{A})$,
- $\cdot\,\,$ bit b_0- whether "left" or "right" messages are processed

- A series of hybrids $\mathcal{B}_{b_0}^{b_1,\dots,b_d}(\mathcal{A})$,
- $\cdot\,\,$ bit b_0- whether "left" or "right" messages are processed
- bit b_i , $i \in \{1, ..., d\}$ what will be used as the *i*-th function: a truly random or pseudorandom function;

- A series of hybrids $\mathcal{B}_{b_0}^{b_1,\dots,b_d}(\mathcal{A})$,
- \cdot bit b_0 whether "left" or "right" messages are processed
- bit b_i , $i \in \{1, ..., d\}$ what will be used as the *i*-th function: a truly random or pseudorandom function;
- process the inputs (m_0, i_0, m_1, i_1) as follows:
 - \cdot if $b_0 = 0$, $b_{i_0} = 0$: return a random string of appropriate length;
 - if $b_0 = 0$, $b_{i_0} = 1$: return $\mathcal{F}_{k_{i_0}}(m_0)$;
 - if $b_0 = 1$, $b_{i_1} = 0$: return a random string of appropriate length;
 - if $b_0 = 1$, $b_{i_0} = 1$: return $\mathcal{F}_{k_{i_1}}(m_1)$;

Introduction

Session key secrecy

Explicit authentication

User privacy

Conclusion

• Three models were analyzed:

- Three models were analyzed:
- PRF⁺: hard to distinguish segments of PRF from a truly random strings in the presence of additional information;

- Three models were analyzed:
- PRF⁺: hard to distinguish segments of PRF from a truly random strings in the presence of additional information;
- UF-PRF: hard to forge segments of PRF;

- Three models were analyzed:
- PRF⁺: hard to distinguish segments of PRF from a truly random strings in the presence of additional information;
- + UF-PRF: hard to forge segments of PRF;
- LOR-PRF: hard to guess which message was processed;

- Three models were analyzed:
- PRF⁺: hard to distinguish segments of PRF from a truly random strings in the presence of additional information;
- UF-PRF: hard to forge segments of PRF;
- LOR-PRF: hard to guess which message was processed;
- Models can be used in the analysis of 5G-AKA protocol security.

- Bellare, Mihir, Oded Goldreich, and Anton Mityagin. The Power of Verification Queries in Message Authentication and Authenticated Encryption. Cryptology ePrint Archive, Paper 2004/309. https://eprint.iacr.org/2004/309. 2004. URL: https://eprint.iacr.org/2004/309.
- Bellare, Mihir, Tadayoshi Kohno, and Chanathip Namprempre. "Breaking and provably repairing the SSH authenticated encryption scheme: A case study of the Encode-then-Encrypt-and-MAC paradigm." In: ACM Transactions on Information and System Security (TISSEC) 7.2 (2004), pp. 206–241.

Thank you for your attention!

Author(s):

Tsaregorodtsev Kirill

Researcher at Cryptography laboratory, JSRPC "Kryptonite", Moscow, Russia k.tsaregorodtsev@kryptonite.ru