On Hiding a Plaintext Length by Preencryption

Cihangir TEZCAN and Serge VAUDENAY

École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

ACNS 2011
June 09, 2011, Nerja, Spain

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Outline

Introduction

- **Problem:** Encryption schemes cannot hide a plaintext length when plaintext domain is unbounded.

Introduction

- **Problem:** Encryption schemes cannot hide a plaintext length when plaintext domain is unbounded. Moreover, an approximation of the plaintext length may leak some information.

Introduction

- **Problem:** Encryption schemes cannot hide a plaintext length when plaintext domain is unbounded. Moreover, an approximation of the plaintext length may leak some information.
- **A Solution:** Use *random padding* before the encryption.

Introduction

- **Problem:** Encryption schemes cannot hide a plaintext length when plaintext domain is unbounded. Moreover, an approximation of the plaintext length may leak some information.
- **A Solution:** Use *random padding* before the encryption.
  - e.g. TLS Protocol version 1.2 allows to pad up to 2^{11} bits to frustrate attacks based on the lengths of exchanged messages (but the resulting length must be a multiple of the block size).

Introduction

- **Problem:** Encryption schemes cannot hide a plaintext length when plaintext domain is unbounded. Moreover, an approximation of the plaintext length may leak some information.
- **A Solution:** Use *random padding* before the encryption.
  - e.g. TLS Protocol version 1.2 allows to pad up to 2^{11} bits to frustrate attacks based on the lengths of exchanged messages (but the resulting length must be a multiple of the block size).
- **Aim:** To formalize preencryption schemes and define appropriate secrecy.

Games and Security

Δ-IND-OTE Game

1. Challenger generates a key K and discloses its public part K_p

Games and Security

Δ-IND-OTE Game

1. Challenger generates a key K and discloses its public part K_p
2. Adversary selects plaintexts x_0 and x_1 where $||x_0| - |x_1|| \leq \Delta$

Games and Security

Δ-IND-OTE Game

1. Challenger generates a key K and discloses its public part K_p
2. Adversary selects plaintexts x_0 and x_1 where $||x_0| - |x_1|| \leq \Delta$
3. Challenger flips a coin b, computes $\mathsf{Enc}_K(x_b) = Y$ and gives Y to the adversary

Games and Security

Δ-IND-OTE Game

1. Challenger generates a key K and discloses its public part K_p
2. Adversary selects plaintexts x_0 and x_1 where $||x_0| - |x_1|| \leq \Delta$
3. Challenger flips a coin b, computes $\text{Enc}_K(x_b) = Y$ and gives Y to the adversary
4. Adversary guesses b' and wins if $b' = b$

Games and Security

Δ-IND-OTE Game

1. Challenger generates a key K and discloses its public part K_p
2. Adversary selects plaintexts x_0 and x_1 where $||x_0| - |x_1|| \leq \Delta$
3. Challenger flips a coin b, computes $\mathrm{Enc}_K(x_b) = Y$ and gives Y to the adversary
4. Adversary guesses b' and wins if $b' = b$

IND-OTE security corresponds to the $\Delta = 0$ case.

Games and Security

Δ-IND-OTE Game

1. Challenger generates a key K and discloses its public part K_p
2. Adversary selects plaintexts x_0 and x_1 where $||x_0| - |x_1|| \leq \Delta$
3. Challenger flips a coin b, computes $\mathrm{Enc}_K(x_b) = Y$ and gives Y to the adversary
4. Adversary guesses b' and wins if $b' = b$

IND-OTE security corresponds to the $\Delta = 0$ case.

Definition

The advantage is $2(\Pr[b = b'] - \frac{1}{2})$. We say that the encryption scheme is Δ-IND-OTE(t, ε)-secure if for all adversary with time complexity limited by t, the advantage is at most ε.

Games and Security

Δ-IND-OTE Game

1. Challenger generates a key K and discloses its public part K_p
2. Adversary selects plaintexts x_0 and x_1 where $||x_0| - |x_1|| \leq \Delta$
3. Challenger flips a coin b, computes $\mathsf{Enc}_K(x_b) = Y$ and gives Y to the adversary
4. Adversary guesses b' and wins if $b' = b$

IND-OTE security corresponds to the $\Delta = 0$ case.

Definition

The advantage is $2(\Pr[b = b'] - \frac{1}{2})$. We say that the encryption scheme is Δ-IND-OTE(t, ε)-secure if for all adversary with time complexity limited by t, the advantage is at most ε.

Something is wrong with this definition (yet the results are provided w.r.t. it).

Games and Security

Δ-IND-OTE Game

1. Challenger generates a key K and discloses its public part K_p
2. Adversary selects plaintexts x_0 and x_1 where $||x_0| - |x_1|| \leq \Delta$
3. Challenger flips a coin b, computes $\mathrm{Enc}_K(x_b) = Y$ and gives Y to the adversary
4. Adversary guesses b' and wins if $b' = b$

IND-OTE security corresponds to the $\Delta = 0$ case.

Definition

The advantage is $\Pr[b = b'] - \frac{1}{2}$. We say that the encryption scheme is Δ-IND-OTE(t, ε)-secure if for all adversary with time complexity limited by t, the advantage is at most ε.

This is the definition that is provided in the paper (and it is valid for this talk).

Preencryption Schemes

Definition

Given two plaintext domains $\mathcal{X}$ and $\mathcal{X}^0$, a preencryption scheme from $\mathcal{X}$ to $\mathcal{X}^0$ is a pair of algorithms

- a (probabilistic) algorithm *pre* such that for all $x \in \mathcal{X}$, $pre(x) \in \mathcal{X}^0$ with probability 1
- a (deterministic) algorithm *Extract*

where $Extract(pre(x)) = x$ with probability 1.

Preencryption Schemes

Definition

Given two plaintext domains $\mathcal{X}$ and $\mathcal{X}^0$, a preencryption scheme from $\mathcal{X}$ to $\mathcal{X}^0$ is a pair of algorithms

- a (probabilistic) algorithm *pre* such that for all $x \in \mathcal{X}$, $pre(x) \in \mathcal{X}^0$ with probability 1
- a (deterministic) algorithm *Extract*

where $Extract(pre(x)) = x$ with probability 1.

- a preencryption scheme is *B-almost length preserving* if $||\text{pre}(x)| - |x|| \leq B$ with probability 1 for all x.

Preencryption Schemes

Definition

Given two plaintext domains $\mathcal{X}$ and $\mathcal{X}^0$, a preencryption scheme from $\mathcal{X}$ to $\mathcal{X}^0$ is a pair of algorithms

- a (probabilistic) algorithm *pre* such that for all $x \in \mathcal{X}$, $pre(x) \in \mathcal{X}^0$ with probability 1
- a (deterministic) algorithm *Extract*

where $Extract(pre(x)) = x$ with probability 1.

- a preencryption scheme is *B-almost length preserving* if $||\mathrm{pre}(x)| - |x|| \leq B$ with probability 1 for all x.
- a preencryption scheme is *length-increasing* if $|\mathrm{pre}(x)| \geq |x|$ with probability 1 for all x.

Preencryption Schemes

Δ-IND Game:

1. Adversary selects plaintexts x_0 and x_1 where $||x_0| - |x_1|| \leq \Delta$

Preencryption Schemes

Δ-IND Game:

1. Adversary selects plaintexts x_0 and x_1 where $||x_0| - |x_1|| \leq \Delta$
2. Challenger flips a coin b, computes $|\text{pre}(x_b)| = L$ and gives L to the adversary

Preencryption Schemes

Δ-IND Game:

1. Adversary selects plaintexts x_0 and x_1 where $||x_0| - |x_1|| \leq \Delta$
2. Challenger flips a coin b, computes $|\text{pre}(x_b)| = L$ and gives L to the adversary
3. Adversary guesses b' and wins if $b' = b$

Preencryption Schemes

Δ-IND Game:

1. Adversary selects plaintexts x_0 and x_1 where $||x_0| - |x_1|| \leq \Delta$
2. Challenger flips a coin b, computes $|\text{pre}(x_b)| = L$ and gives L to the adversary
3. Adversary guesses b' and wins if $b' = b$

Definition (Security and Advantage)

A preencryption scheme is Δ-IND (t, ε)-secure if for all adversary $\mathcal{A}$ with time complexity limited by t, the advantage in the following game is at most ε. The advantage is defined as $\Pr[b = b'] - \frac{1}{2}$.

Preencryption Schemes

Theorem

For an IND-OTE-secure encryption C^0 which fully leaks the plaintext length, the Δ-IND security of P is necessary and sufficient to have C Δ-IND-OTE-secure where $C(x) = C^0(pre(x))$.

Preencryption Schemes

Theorem

For an IND-OTE-secure encryption C^0 which fully leaks the plaintext length, the Δ-IND security of P is necessary and sufficient to have C Δ-IND-OTE-secure where $C(x) = C^0(pre(x))$.

i.e. P Δ-IND-secure + C^0 IND-OTE-secure => C Δ-IND-OTE-secure

Advantage

Definition

Given a set of integers A, x_0 and x_1, we define a Δ-IND adversary $D_A(x_0, x_1)$ as the one selecting x_0 and x_1 then yielding $b' = 1$ if and only if $L \in A$. We define $\mathrm{Adv}_A(x_0, x_1)$ as the advantage of this adversary.

Advantage

Definition

Given a set of integers A, x_0 and x_1, we define a Δ-IND adversary $D_A(x_0, x_1)$ as the one selecting x_0 and x_1 then yielding $b' = 1$ if and only if $L \in A$. We define $\mathsf{Adv}_A(x_0, x_1)$ as the advantage of this adversary.

Notation

We denote $\mathsf{Adv}(x_0, x_1)$ as the maximal advantage for adversaries selecting x_0 and x_1.

Advantage

Definition

Given a set of integers A, x_0 and x_1, we define a Δ-IND adversary $D_A(x_0, x_1)$ as the one selecting x_0 and x_1 then yielding $b' = 1$ if and only if $L \in A$. We define $\mathrm{Adv}_A(x_0, x_1)$ as the advantage of this adversary.

Notation

We denote $\mathrm{Adv}(x_0, x_1)$ as the maximal advantage for adversaries selecting x_0 and x_1.

Actually, $\mathrm{Adv}(x_0, x_1)$ is the statistical distance between $|\mathrm{pre}(x_0)|$ and $|\mathrm{pre}(x_1)|$.

Maximal Security of the Pad-then-Encrypt Scheme

Definition

A padding scheme defines the preencryption scheme $\text{pre}(x) = x \| \text{pad}(x)$.

Maximal Security of the Pad-then-Encrypt Scheme

Definition

A padding scheme defines the preencryption scheme $\mathsf{pre}(x) = x \| \mathsf{pad}(x)$.

Note that preencryption schemes made out from a padding scheme are all length-increasing.

Maximal Security of the Pad-then-Encrypt Scheme

Definition

A padding scheme defines the preencryption scheme $\mathsf{pre}(x) = x \| \mathsf{pad}(x)$.

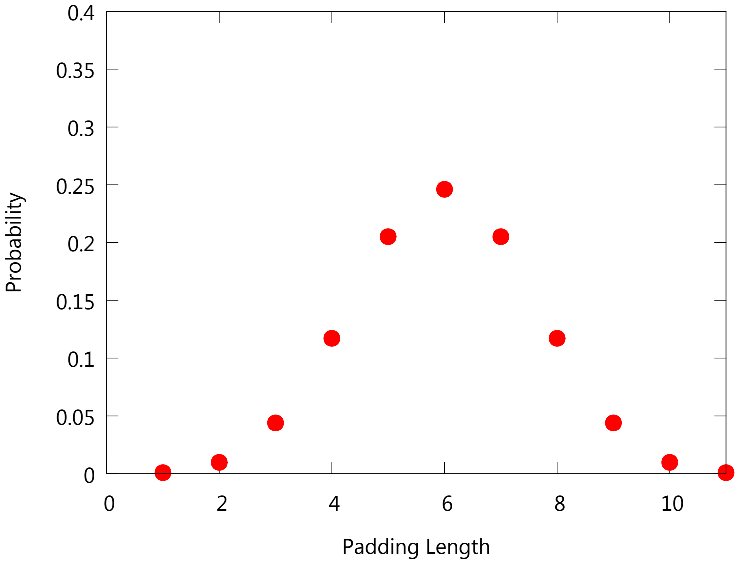
Note that preencryption schemes made out from a padding scheme are all length-increasing.

Example

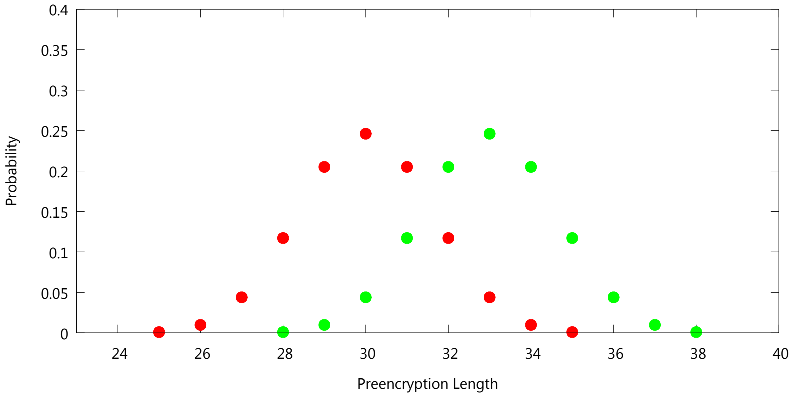
Let $B = 11$ and N be the binomial distribution with parameters 10 and $\frac{1}{2}$.

Let the lengths of the two chosen plaintexts for the Δ-IND game be $|x_0| = 24$ and $|x_1| = 27$.

An Example


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Probability
0
2
4
6
8
10
Padding Length

An Example


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Probability
24
26
28
30
32
34
36
38
40
Preencryption Length

Maximal Security of the Pad-then-Encrypt Scheme

Theorem (Lower bound)

If P is length-increasing and B-almost length-preserving, then there exists an adversary with advantage at least $\frac{1}{2\left\lceil \frac{B}{\Delta} \right\rceil}$.

Maximal Security of the Pad-then-Encrypt Scheme

Theorem (Lower bound)

If P is length-increasing and B-almost length-preserving, then there exists an adversary with advantage at least $\frac{1}{2\left\lceil \frac{B}{\Delta} \right\rceil}$.

Some assumptions:

- (uniformity) the distribution of the padding length is fixed (it does not depend on the plaintext)

Maximal Security of the Pad-then-Encrypt Scheme

Theorem (Lower bound)

If P is length-increasing and B-almost length-preserving, then there exists an adversary with advantage at least $\frac{1}{2\lceil \frac{B}{\Delta} \rceil}$.

Some assumptions:

- (uniformity) the distribution of the padding length is fixed (it does not depend on the plaintext)
- (almost length-preserving) the padding length is in $\{1, \ldots, B\}$

Uniform Padding Schemes

We are considering the Δ-IND game where $||x_0| - |x_1|| \leq \Delta$, N is the distribution for the padding length, and $|pad(x)| \leq B$. Three questions to answer:

1. Given B and Δ, what is the optimal distribution N?

Uniform Padding Schemes

We are considering the Δ-IND game where $||x_0| - |x_1|| \leq \Delta$, N is the distribution for the padding length, and $|pad(x)| \leq B$. Three questions to answer:

1. Given B and Δ, what is the optimal distribution N? (uniform distribution is nearly optimal)

Uniform Padding Schemes

We are considering the Δ-IND game where $||x_0| - |x_1|| \leq \Delta$, N is the distribution for the padding length, and $|pad(x)| \leq B$. Three questions to answer:

1. Given B and Δ, what is the optimal distribution N? (uniform distribution is nearly optimal)
2. What is the ε-security of the optimal distribution?

Uniform Padding Schemes

We are considering the Δ-IND game where $||x_0| - |x_1|| \leq \Delta$, N is the distribution for the padding length, and $|pad(x)| \leq B$. Three questions to answer:

1. Given B and Δ, what is the optimal distribution N? (uniform distribution is nearly optimal)
2. What is the ε-security of the optimal distribution? (nearly $\frac{\Delta}{2B}$)

Uniform Padding Schemes

We are considering the Δ-IND game where $||x_0| - |x_1|| \leq \Delta$, N is the distribution for the padding length, and $|pad(x)| \leq B$. Three questions to answer:

1. Given B and Δ, what is the optimal distribution N? (uniform distribution is nearly optimal)
2. What is the ε-security of the optimal distribution? (nearly $\frac{\Delta}{2B}$)
3. Given Δ, to obtain ε-security, what should be the padding length B?

Uniform Padding Schemes

We are considering the Δ-IND game where $||x_0| - |x_1|| \leq \Delta$, N is the distribution for the padding length, and $|pad(x)| \leq B$. Three questions to answer:

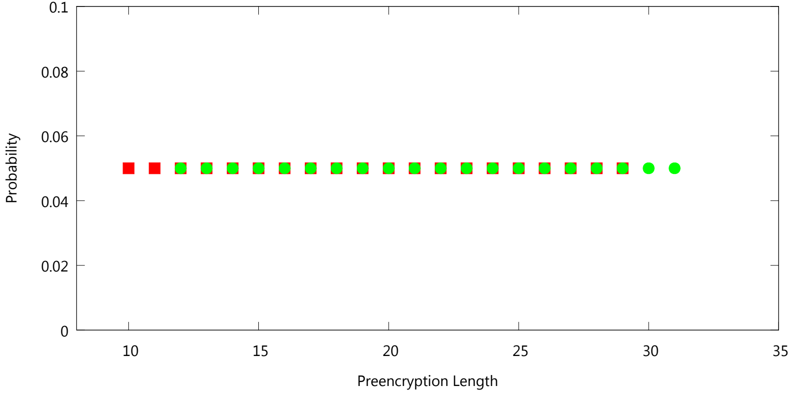
1. Given B and Δ, what is the optimal distribution N? (uniform distribution is nearly optimal)
2. What is the ε-security of the optimal distribution? (nearly $\frac{\Delta}{2B}$)
3. Given Δ, to obtain ε-security, what should be the padding length B? (nearly $\frac{\Delta}{2\varepsilon}$)

Uniform Padding Schemes

Example

The padding scheme that has uniformly distributed padding length in $\{1, \dots, B\}$ has advantage $\mathsf{Adv}(x_0, x_1) = \frac{||x_1|-|x_0||}{2B}$. So, this preencryption scheme is Δ-IND$\left(t, \frac{\Delta}{2B}\right)$-secure for all Δ and any t.

Example: Uniform Distribution



Uniform Padding Schemes

Thus, we have $\frac{\Delta}{2B} \geq \mathrm{Adv}(a, b) \geq \frac{1}{2\left\lceil \frac{B}{\Delta} \right\rceil}$.

Uniform Padding Schemes

Thus, we have $\frac{\Delta}{2B} \geq \text{Adv}(a, b) \geq \frac{1}{2\lceil \frac{B}{\Delta} \rceil}$.

Theorem ($\Delta = 2$ Case)

Consider a uniform strictly length-increasing and B-almost length-preserving padding scheme. If B is odd and $\Delta = 2$ then $\text{Adv}(a, b) \geq \frac{B}{B^2+1}$.

Table: Security when $\Delta = 2$ and B is odd

B	**Uniform Distribution** $\frac{\Delta}{2B}$	**Best Achievable** $\frac{B}{B^2+1}$	**Lower Bound** $\frac{1}{2\lceil \frac{B}{\Delta} \rceil}$
3	0.333333333333333	0.3	0.25
5	0.2	0.192307692307692	0.166666666666667
7	0.142857142857143	0.14	0.125
9	0.111111111111111	0.109756097560976	0.1
11	0.0909090909090909	0.0901639344262295	0.0833333333333333
13	0.0769230769230769	0.0764705882352941	0.0714285714285714
15	0.0666666666666667	0.0663716814159292	0.0625
17	0.0588235294117647	0.0586206896551724	0.0555555555555556
19	0.0526315789473684	0.0524861878453039	0.05
21	0.0476190476190476	0.0475113122171946	0.0454545454545455
23	0.0434782608695652	0.0433962264150943	0.0416666666666667
25	0.04	0.0399361022364217	0.0384615384615385
27	0.037037037037037	0.036986301369863	0.0357142857142857
29	0.0344827586206897	0.0344418052256532	0.0333333333333333
31	0.032258064516129	0.0322245322245322	0.03125
33	0.0303030303030303	0.0302752293577982	0.0294117647058824
35	0.0285714285714286	0.0285481239804241	0.0277777777777778
37	0.027027027027027	0.027007299270073	0.0263157894736842
39	0.0256410256410256	0.0256241787122208	0.025
41	0.024390243902439	0.0243757431629013	0.0238095238095238
43	0.0232558139534884	0.0232432432432432	0.0227272727272727
45	0.0222222222222222	0.0222112537018756	0.0217391304347826
47	0.0212765957446809	0.0212669683257919	0.0208333333333333
49	0.0204081632653061	0.0203996669442132	0.02

Some Consequences

- TLS Protocol version 1.2 allows to pad up to $B = 2^{11}$ bits to frustrate attacks based on the lengths of exchanged messages. So it is Δ-IND$(t, \frac{\Delta}{2^{12}})$-secure.

Some Consequences

- TLS Protocol version 1.2 allows to pad up to $B = 2^{11}$ bits to frustrate attacks based on the lengths of exchanged messages. So it is Δ-IND(t, $\frac{\Delta}{2^{12}}$)-secure. However, the resulting length must be a multiple of the block size. For example, $B = 32$ blocks of data when the block cipher uses blocks of 64 bits. So the real security is $\varepsilon = \frac{\Delta}{2^5}$.

Some Consequences

- TLS Protocol version 1.2 allows to pad up to $B = 2^{11}$ bits to frustrate attacks based on the lengths of exchanged messages. So it is Δ-IND$(t, \frac{\Delta}{2^{12}})$-secure. However, the resulting length must be a multiple of the block size. For example, $B = 32$ blocks of data when the block cipher uses blocks of 64 bits. So the real security is $\varepsilon = \frac{\Delta}{2^5}$.
- Usual security levels cannot be obtained for the Δ-IND-OTE game in practice. e.g. To have 2^{-80}-indistinguishable two plaintexts with a single bit of length difference (i.e. 1-IND-OTE$(t, 2^{-80})$), we need to append a padding of length 2^{79} bits.

Conclusion

- We formalized the notion of preencryption scheme and its associated Δ-IND security notion.

Conclusion

- We formalized the notion of preencryption scheme and its associated Δ-IND security notion.
- We formalized the pad-then-encrypt technique and showed that Δ-IND-security is necessary and sufficient to make an encryption scheme Δ-IND-OTE secure.

Conclusion

- We formalized the notion of preencryption scheme and its associated Δ-IND security notion.
- We formalized the pad-then-encrypt technique and showed that Δ-IND-security is necessary and sufficient to make an encryption scheme Δ-IND-OTE secure.
- We showed that there is always an adversary with advantage nearly $\frac{\Delta}{2B}$. So, insecurity degrades linearly with the padding length B.

Conclusion

- We formalized the notion of preencryption scheme and its associated Δ-IND security notion.
- We formalized the pad-then-encrypt technique and showed that Δ-IND-security is necessary and sufficient to make an encryption scheme Δ-IND-OTE secure.
- We showed that there is always an adversary with advantage nearly $\frac{\Delta}{2B}$. So, insecurity degrades linearly with the padding length B.
- We showed that a padding scheme making padding lengths uniformly distributed is nearly optimal.

Conclusion

THANK YOU FOR YOUR ATTENTION