

RUHR-UNIVERSITÄT BOCHUM

### The Subset-Sum Problem

cryptanalysis employing a probabilistic approach

**© KU Leuven**, 22.09.2016 – 13.30

#### Matthias Minihold

ECRYPT-NET Early Stage Researcher Cryptology and IT-Security, Ruhr-Universität Bochum





- 1 The Subset-Sum problem
  - Motivation
  - Historical Remarks
  - Easy and Hard Instances
  - Evolution of Algorithms
  - Technique 1 Meet in the Middle
  - Technique 2 Enlarge Number Set
  - Applications





#### Outline

- 1 The Subset-Sum problem
  - Motivation
  - Historical Remarks
  - Easy and Hard Instances
  - Evolution of Algorithms
  - Technique 1 Meet in the Middle
  - Technique 2 Enlarge Number Set
  - Applications











Informally: Pack a knapsack with items to meet a weight constraint.





Informally: Pack a knapsack with items to meet a weight constraint. Suppose you are at the airport:

▶ The luggage may weight  $\leq S$  [kg] at check-in.





Informally. Pack a knapsack with items to meet a weight constr

Informally: Pack a knapsack with items to meet a weight constraint. Suppose you are at the airport:

- ▶ The luggage may weight  $\leq S$  [kg] at check-in.
- ▶ Not squandering, the bag WILL weight S [kg].





Informally: Pack a knapsack with items to meet a weight constraint. Suppose you are at the airport:

- ▶ The luggage may weight  $\leq S$  [kg] at check-in.
- ▶ Not squandering, the bag WILL weight *S* [kg].
- ▶ *n* equally beautiful items with weights  $a_1, a_2, ... a_n$ .





Informally: Pack a knapsack with items to meet a weight constraint. Suppose you are at the airport:

- ▶ The luggage may weight  $\leq S$  [kg] at check-in.
- ▶ Not squandering, the bag WILL weight *S* [kg].
- ▶ *n* equally beautiful items with weights  $a_1, a_2, ... a_n$ .





Informally: Pack a knapsack with items to meet a weight constraint. Suppose you are at the airport:

- ▶ The luggage may weight  $\leq S$  [kg] at check-in.
- $\blacktriangleright$  Not squandering, the bag WILL weight S [kg].
- ▶ n equally beautiful items with weights  $a_1, a_2, \ldots a_n$ .



## Definition 1 (Subset-Sum)

Given 
$$n, S, a_1, a_2, \dots a_n \in \mathbb{N}$$
, find  $I \subseteq [n] : \sum_{i \in I} a_i = S$ . (1)



#### Outline

- 1 The Subset-Sum problem
  - Motivation
  - Historical Remarks
  - Easy and Hard Instances
  - Evolution of Algorithms
  - Technique 1 Meet in the Middle
  - Technique 2 Enlarge Number Set
  - Applications





▶ Names: (0-1-)Knapsack, Subset-Sum problem



- ► Names: (0-1-)Knapsack, Subset-Sum problem
- ► First studied in 1897



- ▶ Names: (0-1-)Knapsack, Subset-Sum problem
- ► First studied in 1897
- $\blacktriangleright$  Early recognized to be  $\mathcal{NP}\text{--complete}$  by reduction to 3-SAT:



- ▶ Names: (0-1-)Knapsack, Subset-Sum problem
- First studied in 1897
- $\blacktriangleright$  Early recognized to be  $\mathcal{NP}$ -complete by reduction to 3-SAT:
  - $\circ$  Satisfiability with at most 3 literals per clause ( $\hat{=}$  3-SAT)





- ► Names: (0-1-)Knapsack, Subset-Sum problem
- First studied in 1897
- ▶ Early recognized to be NP-complete by reduction to 3-SAT:
  - ∘ Satisfiability with at most 3 literals per clause (≜ 3-SAT)
  - $\circ \Rightarrow_R$  Chromatic number (= Graph Coloring)



- ► Names: (0-1-)Knapsack, Subset-Sum problem
- First studied in 1897
- ▶ Early recognized to be  $\mathcal{NP}$ -complete by reduction to 3-SAT:
  - Satisfiability with at most 3 literals per clause (= 3-SAT)
  - $\circ \Rightarrow_R$  Chromatic number (= Graph Coloring)
  - $\circ \Rightarrow_R \mathsf{Exact} \mathsf{cover}$





- ► Names: (0-1-)Knapsack, Subset-Sum problem
- First studied in 1897
- ▶ Early recognized to be  $\mathcal{NP}$ -complete by reduction to 3-SAT:
  - Satisfiability with at most 3 literals per clause (\(\hat{=}\) 3-SAT)
  - $\circ \Rightarrow_R$  Chromatic number (= Graph Coloring)
  - $\circ \Rightarrow_R \mathsf{Exact} \mathsf{cover}$
  - $\circ \Rightarrow_R \mathsf{Knapsack}$





- ► Names: (0-1-)Knapsack, Subset-Sum problem
- First studied in 1897
- ▶ Early recognized to be  $\mathcal{NP}$ -complete by reduction to 3-SAT:
  - Satisfiability with at most 3 literals per clause (≜ 3-SAT)
  - $\circ \Rightarrow_R$  Chromatic number (= Graph Coloring)
  - $\circ \Rightarrow_R \mathsf{Exact} \mathsf{cover}$
  - $\circ \Rightarrow_R \mathsf{Knapsack}$
  - worst-case instances are computationally intractable



#### Outline

- 1 The Subset-Sum problem
  - Motivation
  - Historical Remarks
  - Easy and Hard Instances
  - Evolution of Algorithms
  - Technique 1 Meet in the Middle
  - Technique 2 Enlarge Number Set
  - Applications







Given an instance by  $n, S, a_1, a_2, \dots a_n \in \mathbb{N}$  with density  $d := \frac{n}{\log(\max_i a_i)}$ .



Given an instance by 
$$n, S, a_1, a_2, \dots a_n \in \mathbb{N}$$
 with density  $d := \frac{n}{\log(\max_i a_i)}$ . Easy instances:

▶ (d<0.9408) Solve Shortest Vector Problem in  $\mathcal{L}(B)$ .



Given an instance by 
$$n, S, a_1, a_2, \ldots a_n \in \mathbb{N}$$
 with density  $d := \frac{n}{\log(\max_i a_i)}$ . Easy instances:

- ▶ (d<0.9408) Solve Shortest Vector Problem in  $\mathcal{L}(B)$ .
- ▶ (d>1) Solve Dynamic Programming Problem



Given an instance by 
$$n, S, a_1, a_2, \ldots a_n \in \mathbb{N}$$
 with density  $d := \frac{n}{\log(\max_i a_i)}$ . Easy instances:

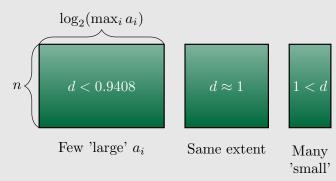
- ▶ (d<0.9408) Solve Shortest Vector Problem in  $\mathcal{L}(B)$ .
- ▶ (d>1) Solve Dynamic Programming Problem





Given an instance by  $n, S, a_1, a_2, \dots a_n \in \mathbb{N}$  with density  $d := \frac{n}{\log(\max_i a_i)}$ . Easy instances:

- ▶ (d<0.9408) Solve Shortest Vector Problem in  $\mathcal{L}(B)$ .
- ▶ (d>1) Solve Dynamic Programming Problem



 $a_i$ 



#### Hard instances:

▶ Element size:  $\log_2 a_i \approx n$  for i = 1, 2, ..., n



- ▶ Element size:  $\log_2 a_i \approx n$  for i = 1, 2, ..., n
- ▶ Characterized by  $0.9408 < d = \frac{n}{\log(\max_i a_i)} < 1$



- ▶ Element size:  $\log_2 a_i \approx n$  for i = 1, 2, ..., n
- ▶ Characterized by  $0.9408 < d = \frac{n}{\log(\max_i a_i)} < 1$
- ▶ Balanced solution:  $|I| \approx \frac{n}{2}$



- ▶ Element size:  $\log_2 a_i \approx n$  for i = 1, 2, ..., n
- ► Characterized by  $0.9408 < d = \frac{n}{\log(\max_i a_i)} < 1$
- ▶ Balanced solution:  $|I| \approx \frac{n}{2}$
- ▶ How to generate hard instances parametrized by n:



- ▶ Element size:  $\log_2 a_i \approx n$  for i = 1, 2, ..., n
- ► Characterized by  $0.9408 < d = \frac{n}{\log(\max_i a_i)} < 1$
- ▶ Balanced solution:  $|I| \approx \frac{n}{2}$
- ▶ How to generate hard instances parametrized by *n*:
  - 1. Choose  $a \in_R \{0, 1, \dots, 2^n 1\}^n, x \in_R \{0, 1\}^n$  uniformly.

- ▶ Element size:  $\log_2 a_i \approx n$  for i = 1, 2, ..., n
- ► Characterized by  $0.9408 < d = \frac{n}{\log(\max_i a_i)} < 1$
- ▶ Balanced solution:  $|I| \approx \frac{n}{2}$
- ▶ How to generate hard instances parametrized by *n*:
  - 1. Choose  $a \in_R \{0, 1, \dots, 2^n 1\}^n, x \in_R \{0, 1\}^n$  uniformly.
  - 2. Output  $(a = (a_1, a_2, ..., a_n), S = a \cdot x = \sum_{i=1}^n x_i a_i)$ .



- ▶ Element size:  $\log_2 a_i \approx n$  for i = 1, 2, ..., n
- ► Characterized by  $0.9408 < d = \frac{n}{\log(\max_i a_i)} < 1$
- ▶ Balanced solution:  $|I| \approx \frac{n}{2}$
- ▶ How to generate hard instances parametrized by *n*:
  - 1. Choose  $a \in_R \{0, 1, \dots, 2^n 1\}^n, x \in_R \{0, 1\}^n$  uniformly.
  - 2. Output  $(a = (a_1, a_2, \dots a_n), S = a \cdot x = \sum_{i=1}^n x_i a_i)$ .
  - 3. Finding x given (a, S) is (assumed to be) computationally hard.



#### Outline

- 1 The Subset-Sum problem
  - Motivation
  - Historical Remarks
  - Easy and Hard Instances
  - Evolution of Algorithms
  - Technique 1 Meet in the Middle
  - Technique 2 Enlarge Number Set
  - Applications





# **Evolution of Algorithms (giving exact solutions)**

Horst Görtz Institut ■ für IT-Sicherheit ■

Algorithm (year)	Time	Space

Table: Expected time and space requirements of algorithms solving Eq. (1).



# **Evolution of Algorithms (giving exact solutions)**



Algorithm (year)	Time	Space
Exhaustive Search	$2^{1.000n} \approx (2.000)^n$	$2^{0.000n} \approx (1.000)^n$

Table: Expected time and space requirements of algorithms solving Eq. (1).



# **Evolution of Algorithms (giving exact solutions)**

Algorithm (year)	Time	Space
Exhaustive Search	$2^{1.000n} \approx (2.000)^n$	$2^{0.000n} \approx (1.000)^n$

> Idea: 'meet-in-the-middle'

Table: Expected time and space requirements of algorithms solving Eq. (1).

Algorithm (year)	Time	Space
Exhaustive Search	$2^{1.000n} \approx (2.000)^n$	$2^{0.000n} \approx (1.000)^n$
> Idea: <i>'meet-in-the-middle'</i> Horowitz-Sahni (1974)	$2^{0.500n} \approx (1.414)^n$	$2^{0.500n} \approx (1.414)^n$







Algorithm (year)	Time	Space
Exhaustive Search	$2^{1.000n} \approx (2.000)^n$	$2^{0.000n} \approx (1.000)^n$
> Idea: <i>'meet-in-the-middle'</i> Horowitz-Sahni (1974)	$2^{0.500n} \approx (1.414)^n$	$2^{0.500n} \approx (1.414)^n$
> Idea: '4-list approach'	·	



hgi
Horst Görtz Institut ■
für IT-Sicherheit ■

Algorithm (year)	Time	Space
Exhaustive Search	$2^{1.000n} \approx (2.000)^n$	$2^{0.000n} \approx (1.000)^n$
> Idea: 'meet-in-the-middle' Horowitz-Sahni (1974) > Idea: '4-list approach'	$2^{0.500n} \approx (1.414)^n$ $2^{0.500n} \approx (1.414)^n$	,
Schroeppel-Shamir (1979)	2000000 ≈ (1.414)00	$2^{0.233.0} \approx (1.189)^{0.0}$





Algorithm (year)	Time	Space
Exhaustive Search	$2^{1.000n} \approx (2.000)^n$	$2^{0.000n} \approx (1.000)^n$
> Idea: 'meet-in-the-middle'		
Horowitz-Sahni (1974)	$2^{0.500n} \approx (1.414)^n$	$2^{0.500n} \approx (1.414)^n$
> Idea: '4-list approach'		
Schroeppel-Shamir (1979)	$2^{0.500n} \approx (1.414)^n$	$2^{0.250n} \approx (1.189)^n$
> Idea: 'multiple representations'		· · ·



Algorithm (year)	Time	Space
Exhaustive Search	$2^{1.000n} \approx (2.000)^n$	$2^{0.000n} \approx (1.000)^n$
> Idea: 'meet-in-the-middle'		
Horowitz-Sahni (1974)	$2^{0.500n} \approx (1.414)^n$	$2^{0.500n} \approx (1.414)^n$
> Idea: '4-list approach'		
Schroeppel-Shamir (1979)	$2^{0.500n} \approx (1.414)^n$	$2^{0.250n} \approx (1.189)^n$
> Idea: 'multiple representations'		
Howgrave-Graham-Joux (2010)	$2^{0.337n} \approx (1.263)^n$	$2^{0.311n} \approx (1.241)^n$
> Idea: 'enlarged number-set' {-1 0 1}		





Algorithm (year)	Time	Space
Exhaustive Search	$2^{1.000n} \approx (2.000)^n$	$2^{0.000n} \approx (1.000)^n$
> Idea: 'meet-in-the-middle'		
Horowitz-Sahni (1974)	$2^{0.500n} \approx (1.414)^n$	$2^{0.500n} \approx (1.414)^n$
> Idea: '4-list approach'		
Schroeppel-Shamir (1979)	$2^{0.500n} \approx (1.414)^n$	$2^{0.250n} \approx (1.189)^n$
> Idea: 'multiple representations'		
Howgrave-Graham-Joux (2010)	$2^{0.337n} \approx (1.263)^n$	$2^{0.311n} \approx (1.241)^n$
$>$ Idea: 'enlarged number-set' $\{-1,0,1\}$		
Becker-Coron-Joux (2011)	$2^{0.291n} \approx (1.223)^n$	$2^{0.291n} \approx (1.223)^n$
` '	,	,





Algorithm (year)	Time	Space
Exhaustive Search	$2^{1.000n} \approx (2.000)^n$	$2^{0.000n} \approx (1.000)^n$
> Idea: 'meet-in-the-middle'		
Horowitz-Sahni (1974)	$2^{0.500n} \approx (1.414)^n$	$2^{0.500n} \approx (1.414)^n$
> Idea: '4-list approach'		
Schroeppel-Shamir (1979)	$2^{0.500n} \approx (1.414)^n$	$2^{0.250n} \approx (1.189)^n$
> Idea: 'multiple representations'		
Howgrave-Graham-Joux (2010)	$2^{0.337n} \approx (1.263)^n$	$2^{0.311n} \approx (1.241)^n$
$>$ Idea: 'enlarged number-set' $\{-1,0,1\}$		
Becker-Coron-Joux (2011)	$2^{0.291n} \approx (1.223)^n$	$2^{0.291n} \approx (1.223)^n$
> Idea: 'quantum algorithm'		



Algorithm (year)	Time	Space
Exhaustive Search	$2^{1.000n} \approx (2.000)^n$	$2^{0.000n} \approx (1.000)^n$
> Idea: 'meet-in-the-middle'		
Horowitz-Sahni (1974)	$2^{0.500n} \approx (1.414)^n$	$2^{0.500n} \approx (1.414)^n$
> Idea: '4-list approach'		
Schroeppel-Shamir (1979)	$2^{0.500n} \approx (1.414)^n$	$2^{0.250n} \approx (1.189)^n$
> Idea: 'multiple representations'		
Howgrave-Graham-Joux (2010)	$2^{0.337n} \approx (1.263)^n$	$2^{0.311n} \approx (1.241)^n$
$>$ Idea: 'enlarged number-set' $\{-1,0,1\}$		
Becker-Coron-Joux (2011)	$2^{0.291n} \approx (1.223)^n$	$2^{0.291n} \approx (1.223)^n$
> Idea: 'quantum algorithm'		
Bernstein-Jeffery-Lange-Meurer (2013)	$2^{0.241n} \approx (1.182)^n$	$2^{0.241n} \approx (1.182)^n$

Table: Expected time and space requirements of algorithms solving Eq. (1).

Asymptotically best exact algorithm is a quantum algorithm



hgı	
Horst Görtz Institut ■	
für IT-Sicherheit	

Algorithm (year)	Time	Space
Exhaustive Search	$2^{1.000n} \approx (2.000)^n$	$2^{0.000n} \approx (1.000)^n$
> Idea: 'meet-in-the-middle'		
Horowitz-Sahni (1974)	$2^{0.500n} \approx (1.414)^n$	$2^{0.500n} \approx (1.414)^n$
> Idea: '4-list approach'		
Schroeppel-Shamir (1979)	$2^{0.500n} \approx (1.414)^n$	$2^{0.250n} \approx (1.189)^n$
> Idea: 'multiple representations'		
Howgrave-Graham-Joux (2010)	$2^{0.337n} \approx (1.263)^n$	$2^{0.311n} \approx (1.241)^n$
$>$ Idea: 'enlarged number-set' $\{-1,0,1\}$		
Becker-Coron-Joux (2011)	$2^{0.291n} \approx (1.223)^n$	$2^{0.291n} \approx (1.223)^n$
> Idea: 'quantum algorithm'		
Bernstein-Jeffery-Lange-Meurer (2013)	$2^{0.241n} \approx (1.182)^n$	$2^{0.241n} \approx (1.182)^n$

- Asymptotically best exact algorithm is a quantum algorithm
- ► Lower bound on running time is unkown

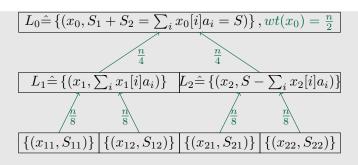


#### Outline

- 1 The Subset-Sum problem
  - Motivation
  - Historical Remarks
  - Easy and Hard Instances
  - Evolution of Algorithms
  - Technique 1 Meet in the Middle
  - Technique 2 Enlarge Number Set
  - Applications









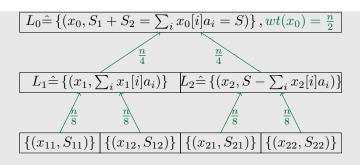
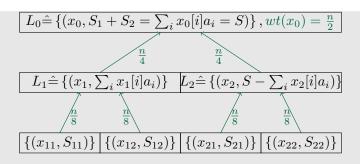


Figure: Schröppel-Shamir: Combining disjoint sub-problems of smaller weight.

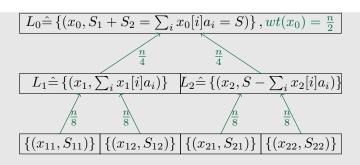
▶ Identify subsets  $I \subseteq [n]$  with vectors  $x \in \{0,1\}^n : i \in I \Leftrightarrow x[i] = 1$ 





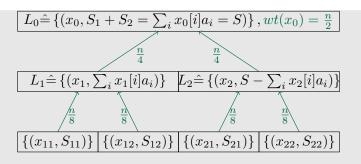
- ▶ Identify subsets  $I \subseteq [n]$  with vectors  $x \in \{0,1\}^n : i \in I \Leftrightarrow x[i] = 1$
- Construct collisions based on the birthday-paradox





- ▶ Identify subsets  $I \subseteq [n]$  with vectors  $x \in \{0,1\}^n : i \in I \Leftrightarrow x[i] = 1$
- Construct collisions based on the birthday-paradox
- $\blacktriangleright$  Construct lists  $L_1, L_2$  of (vector, sum)-pairs in smaller dimension





- ▶ Identify subsets  $I \subseteq [n]$  with vectors  $x \in \{0,1\}^n : i \in I \Leftrightarrow x[i] = 1$
- Construct collisions based on the birthday-paradox
- ightharpoonup Construct lists  $L_1, L_2$  of (vector, sum)-pairs in smaller dimension
- Subset-Sum Merge  $U_{2}$  to solutions  $(x_0,S)\in L_0$  of the Subset-Sum problem  $U_{3/18}$



#### Outline

- 1 The Subset-Sum problem
  - Motivation
  - Historical Remarks
  - Easy and Hard Instances
  - Evolution of Algorithms
  - Technique 1 Meet in the Middle
  - Technique 2 Enlarge Number Set
  - Applications





$$L_0 = \{(x_0, \sum_i x_0[i]a_i = S_1 + S_2 = S)\}, wt(x_0) = \frac{n}{2}\}$$

$$=$$

$$L_1 = \{(x_1, \sum_i x_1[i]a_i)\}, wt(x_1) = \frac{n}{4}$$

$$+$$

$$L_2 = \{(x_2, S - \sum_i x_2[i]a_i)\}, wt(x_2) = \frac{n}{4}$$

Figure: Adding length n solutions of sub-problems enlarges the number-set.



$$L_0 = \{(x_0, \sum_i x_0[i]a_i = S_1 + S_2 = S)\}, wt(x_0) = \frac{n}{2}$$

$$=$$

$$L_1 = \{(x_1, \sum_i x_1[i]a_i)\}, wt(x_1) = \frac{n}{4}$$
 +

$$L_2 = \{(x_2, S - \sum_i x_2[i]a_i)\}, wt(x_2) = \frac{n}{4}$$

Figure: Adding length n solutions of sub-problems enlarges the number-set.

▶ Merge non-disjoint partial solutions in  $L_1, L_2$  filtering 'inconsistent'



Figure: Adding length n solutions of sub-problems enlarges the number-set.

- $\blacktriangleright$  Merge non-disjoint partial solutions in  $L_1, L_2$  filtering 'inconsistent'
- ▶ Enlarge intermediate 'number-set':  $x_{11}[i] + x_{12}[i] =: x_1[i] \notin \{0, 1\}.$

Figure: Adding length *n* solutions of sub-problems enlarges the number-set.

- lacktriangle Merge non-disjoint partial solutions in  $L_1, L_2$  filtering 'inconsistent'
- ▶ Enlarge intermediate 'number-set':  $x_{11}[i] + x_{12}[i] =: x_1[i] \notin \{0,1\}.$

Subset Supple described in Speed of the supple of the sup



#### Outline

- 1 The Subset-Sum problem
  - Motivation
  - Historical Remarks
  - Easy and Hard Instances
  - Evolution of Algorithms
  - Technique 1 Meet in the Middle
  - Technique 2 Enlarge Number Set
  - Applications





### **Applications**



▶ These cryptanalytic methods are meta-techniques





## **Applications**

- ▶ These cryptanalytic methods are meta-techniques
- lacktriangle Applicable to lattice- or code-based  $\mathcal{NP} ext{-complete}$  problems





- ▶ These cryptanalytic methods are meta-techniques
- $\blacktriangleright$  Applicable to lattice- or code-based  $\mathcal{NP}$ -complete problems







- ▶ These cryptanalytic methods are meta-techniques
- lacktriangle Applicable to lattice- or code-based  $\mathcal{NP} ext{-complete}$  problems

Setup: A preprocessed length n text T is stored in the cloud.







- ▶ These cryptanalytic methods are meta-techniques
- lacktriangle Applicable to lattice- or code-based  $\mathcal{NP} ext{-complete}$  problems

Setup: A preprocessed length n text T is stored in the cloud.

Given: A pattern P of length  $m \le n$ , find all locations of P in T.





- ► These cryptanalytic methods are meta-techniques
- $\blacktriangleright$  Applicable to lattice- or code-based  $\mathcal{NP}$ -complete problems

Setup: A preprocessed length n text T is stored in the cloud.

Given: A pattern P of length  $m \le n$ , find all locations of P in T.

Goal: Cloud learns no information about text T and pattern P.



S. Faust, C. Hazay, and D. Venturi.

Outsourced pattern matching.

Cryptology ePrint Archive, Report 2014/662, 2014.

http://eprint.iacr.org/2014/662.



**RUHR-UNIVERSITÄT** BOCHUM

QUESTIONS?

Thank you for your attention!

