CSE 633: Parallel Algorithms Fall 2012

Parallelized Hash Collision Attacking

Benedikt Budig

Course Instructor: Russ Miller



image source: shutterstock



A hash function is a total function $h : \{0,1\}^* \to \{0,1\}^n$



A hash function is a total function $h : \{0,1\}^* \to \{0,1\}^n$ that maps arbitrarily long strings to strings of a fixed length n.



A hash function is a total function $h : \{0,1\}^* \to \{0,1\}^n$ that maps arbitrarily long strings to strings of a fixed length n.

Project Goal: Find Hash Collisions for given Hash



A hash function is a total function $h : \{0,1\}^* \to \{0,1\}^n$ that maps arbitrarily long strings to strings of a fixed length n.

Project Goal: Find Hash Collisions for given Hash A *hash collision* occurs for two strings x, y if h(x) = h(y),



A hash function is a total function $h : \{0,1\}^* \to \{0,1\}^n$ that maps arbitrarily long strings to strings of a fixed length n.

Project Goal: Find Hash Collisions for given Hash A *hash collision* occurs for two strings x, y if h(x) = h(y), that is, if h maps the two strings to the same hash value.



A hash function is a total function $h : \{0,1\}^* \to \{0,1\}^n$ that maps arbitrarily long strings to strings of a fixed length n.

Project Goal: Find Hash Collisions for given Hash A hash collision occurs for two strings x, y if h(x) = h(y), that is, if h maps the two strings to the same hash value. \longrightarrow find a string y such that h(y) collides with given h(x)

image source: shutterstock



A hash function is a total function $h : \{0,1\}^* \to \{0,1\}^n$ that maps arbitrarily long strings to strings of a fixed length n.

Project Goal: Find Hash Collisions for given Hash A hash collision occurs for two strings x, y if h(x) = h(y), that is, if h maps the two strings to the same hash value. \longrightarrow find a string y such that h(y) collides with given h(x)

Reason to do that?

image source: shutterstock



A hash function is a total function $h : \{0,1\}^* \to \{0,1\}^n$ that maps arbitrarily long strings to strings of a fixed length n.

Project Goal: Find Hash Collisions for given Hash A hash collision occurs for two strings x, y if h(x) = h(y), that is, if h maps the two strings to the same hash value. \longrightarrow find a string y such that h(y) collides with given h(x)

Reason to do that: Cryptographic Applications





A hash function is a total function $h : \{0,1\}^* \to \{0,1\}^n$ that maps arbitrarily long strings to strings of a fixed length n.

Project Goal: Find Hash Collisions for given Hash A hash collision occurs for two strings x, y if h(x) = h(y), that is, if h maps the two strings to the same hash value. \longrightarrow find a string y such that h(y) collides with given h(x)

Reason to do that: Cryptographic Applications

• secure storage of passwords





A hash function is a total function $h : \{0,1\}^* \to \{0,1\}^n$ that maps arbitrarily long strings to strings of a fixed length n.

Project Goal: Find Hash Collisions for given Hash A hash collision occurs for two strings x, y if h(x) = h(y), that is, if h maps the two strings to the same hash value. \longrightarrow find a string y such that h(y) collides with given h(x)

Reason to do that: Cryptographic Applications

- secure storage of passwords
- digital signature schemes

222 C 40

image source: shutterstock



A hash function is a total function $h : \{0,1\}^* \to \{0,1\}^n$ that maps arbitrarily long strings to strings of a fixed length n.

Project Goal: Find Hash Collisions for given Hash A hash collision occurs for two strings x, y if h(x) = h(y), that is, if h maps the two strings to the same hash value. \longrightarrow find a string y such that h(y) collides with given h(x)

Reason to do that: Cryptographic Applications

- secure storage of passwords
- digital signature schemes

We focus on — MD5



image source: shutterstock

Input:

Input: hash md5(x) of unknown string x of length $|x| \le n$

Input: hash md5(x) of unknown string x of length $|x| \le n$ Output:

Input: hash md5(x) of unknown string x of length $|x| \le n$

Output: string y such that md5(y) = md5(x) and $|y| \le n$

Input: hash md5(x) of unknown string x of length $|x| \le n$

Output: string y such that md5(y) = md5(x) and $|y| \le n$

Input: hash md5(x) of unknown string x of length $|x| \le n$

Output: string y such that md5(y) = md5(x) and $|y| \le n$

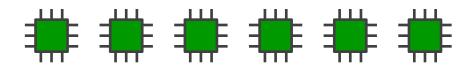
Parallel approach

1. based on their ID, m parallel processes take a subset of the possible strings $\{0,1\}^{\leq n}$

Input: hash md5(x) of unknown string x of length $|x| \le n$ Output: string y such that md5(y) = md5(x) and $|y| \le n$

Parallel approach

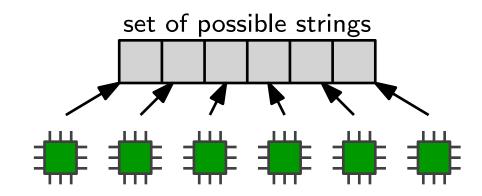
1. based on their ID, m parallel processes take a subset of the possible strings $\{0,1\}^{\leq n}$



Input: hash md5(x) of unknown string x of length $|x| \le n$ Output: string y such that md5(y) = md5(x) and $|y| \le n$

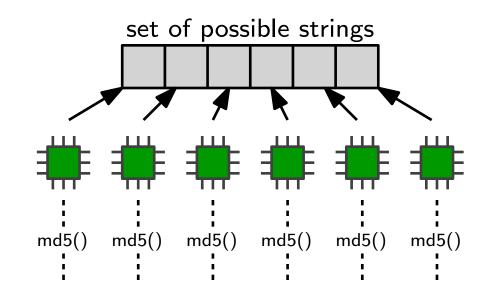
Parallel approach

1. based on their ID, m parallel processes take a subset of the possible strings $\{0,1\}^{\leq n}$



Input: hash md5(x) of unknown string x of length $|x| \le n$ Output: string y such that md5(y) = md5(x) and $|y| \le n$

- 1. based on their ID, m parallel processes take a subset of the possible strings $\{0,1\}^{\leq n}$
- 2. each process calculates hashes of the strings assigned to it



Input: hash md5(x) of unknown string x of length $|x| \le n$ Output: string y such that md5(y) = md5(x) and $|y| \le n$

set of possible strings

⊧ # # # # #

md5() md5() md5() md5() md5()

i

1

md5()

- 1. based on their ID, m parallel processes take a subset of the possible strings $\{0,1\}^{\leq n}$
- 2. each process calculates hashes of the strings assigned to it
- 3. as soon as one process calculates a hash equal to the input hash, all processes terminate

Input: hash md5(x) of unknown string x of length $|x| \le n$ Output: string y such that md5(y) = md5(x) and $|y| \le n$

set of possible strings

iii: #ii: #ii: #ii: #ii

md5() md5() md5() md5() md5()

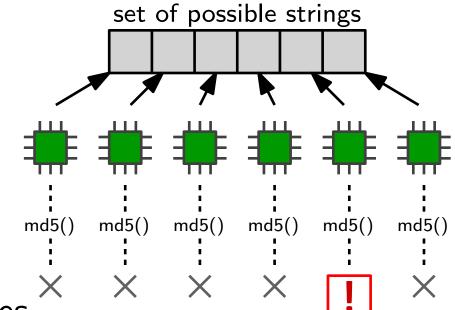
i.

md5()

- 1. based on their ID, m parallel processes take a subset of the possible strings $\{0,1\}^{\leq n}$
- 2. each process calculates hashes of the strings assigned to it
- 3. as soon as one process calculates a hash equal to the input hash, all processes terminate

Input: hash md5(x) of unknown string x of length $|x| \le n$ Output: string y such that md5(y) = md5(x) and $|y| \le n$

- 1. based on their ID, m parallel processes take a subset of the possible strings $\{0,1\}^{\leq n}$
- 2. each process calculates hashes of the strings assigned to it



- 3. as soon as one process calculates a hash equal to the input hash, all processes terminate
- 4. return the colliding string

Hardware



Hardware

• use of CPUs



Hardware

- use of CPUs
- use of a multi-core system



Hardware

- use of CPUs
- use of a multi-core system
- Infiniband network



Hardware

- use of CPUs
- use of a multi-core system
- Infiniband network



image source: CCR website

• tests on CCR machines with 12 cores and 32 cores

Hardware

- use of CPUs
- use of a multi-core system
- Infiniband network



image source: CCR website

• tests on CCR machines with 12 cores and 32 cores

Software and Implementation

Hardware

- use of CPUs
- use of a multi-core system
- Infiniband network



image source: CCR website

• tests on CCR machines with 12 cores and 32 cores

Software and Implementation

• implementation using the C++ programming language

Hardware

- use of CPUs
- use of a multi-core system
- Infiniband network



image source: CCR website

• tests on CCR machines with 12 cores and 32 cores

Software and Implementation

- implementation using the C++ programming language
- one implementation using OpenMP

Hardware

- use of CPUs
- use of a multi-core system
- Infiniband network



image source: CCR website

• tests on CCR machines with 12 cores and 32 cores

Software and Implementation

- implementation using the C++ programming language
- one implementation using OpenMP
- and another version using MPI

Technical Realization

Hardware

- use of CPUs
- use of a multi-core system
- Infiniband network



image source: CCR website

• tests on CCR machines with 12 cores and 32 cores

Software and Implementation

- implementation using the C++ programming language
- one implementation using OpenMP
- and another version using MPI

Future Work

Technical Realization

Hardware

- use of CPUs
- use of a multi-core system
- Infiniband network



image source: CCR website

• tests on CCR machines with 12 cores and 32 cores

Software and Implementation

- implementation using the C++ programming language
- one implementation using OpenMP
- and another version using MPI

Future Work

• tweaks for MPI and OpenMPI

Technical Realization

Hardware

- use of CPUs
- use of a multi-core system
- Infiniband network



image source: CCR website

• tests on CCR machines with 12 cores and 32 cores

Software and Implementation

- implementation using the C++ programming language
- one implementation using OpenMP
- and another version using MPI

Future Work

- tweaks for MPI and OpenMPI
- implementation using CUDA





First Test: OpenMP on 12 Core System



First Test: OpenMP on 12 Core System

• 12 Intel Xeon E5645 at 2.40GHz



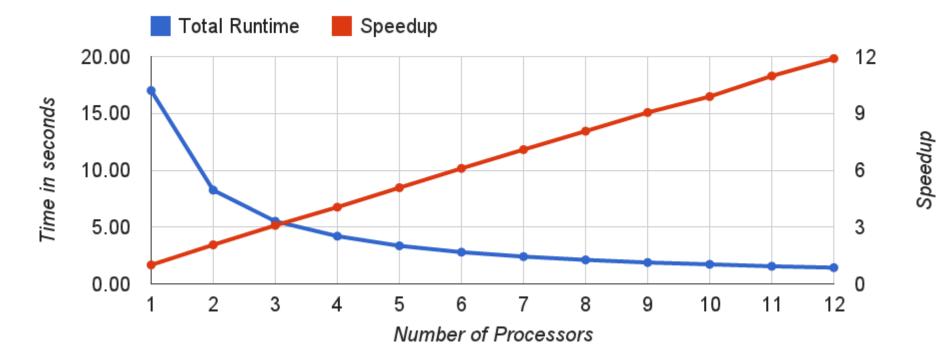
First Test: OpenMP on 12 Core System

- 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0,1\}^{24}$



First Test: OpenMP on 12 Core System

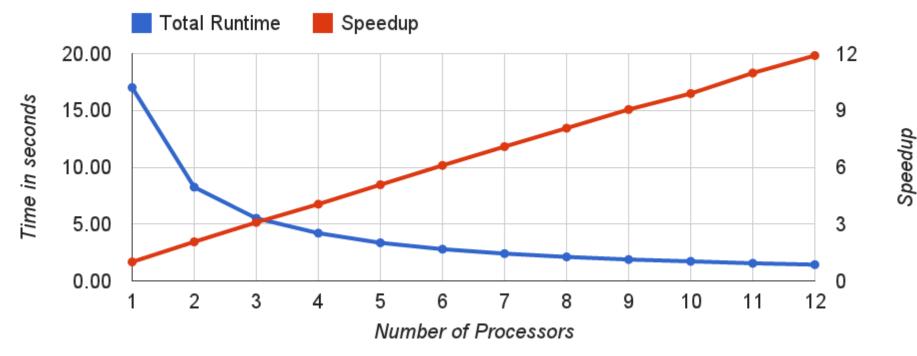
- 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{24}$





First Test: OpenMP on 12 Core System

- 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{24}$

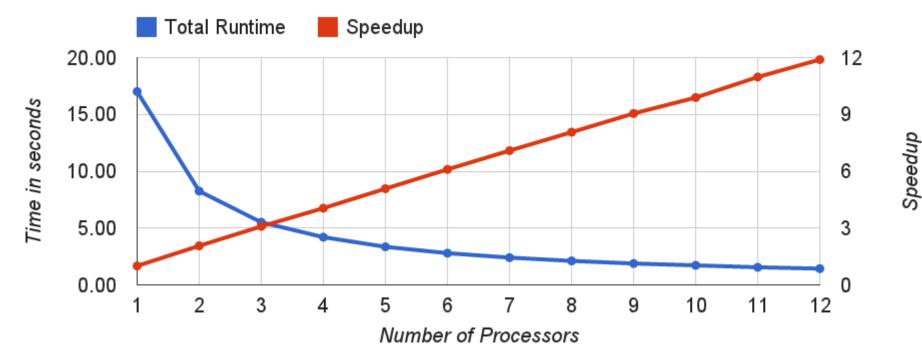


• 1.43 seconds to find collision



First Test: OpenMP on 12 Core System

- 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{24}$

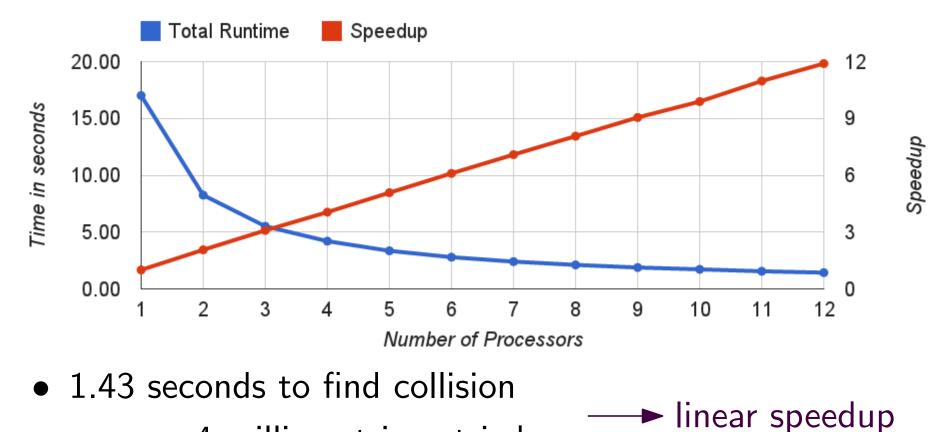


- 1.43 seconds to find collision
- approx. 4 million strings tried



First Test: OpenMP on 12 Core System

- 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{24}$



• approx. 4 million strings tried







Second Test: OpenMP on 12 Core System



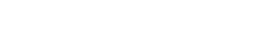
10 - 00 10

Second Test: OpenMP on 12 Core System

• 12 Intel Xeon E5645 at 2.40GHz

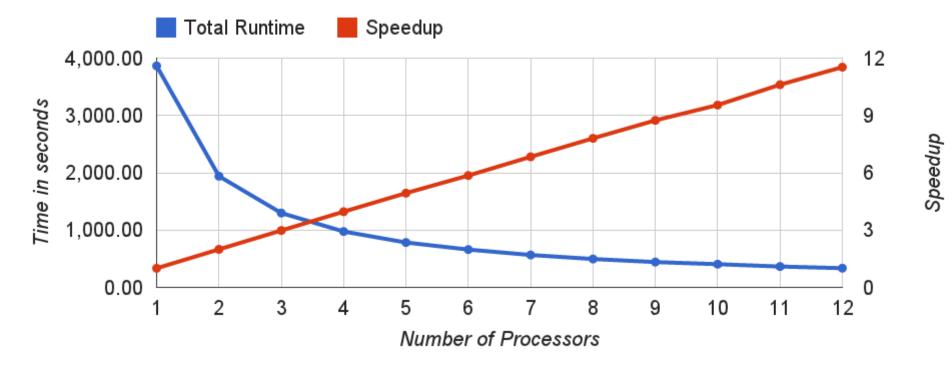
Second Test: OpenMP on 12 Core System

- 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{32}$



Second Test: OpenMP on 12 Core System

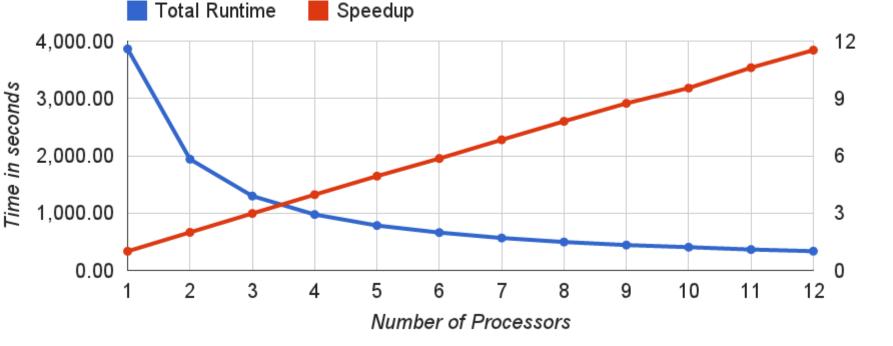
- 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{32}$





Second Test: OpenMP on 12 Core System

- 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{32}$

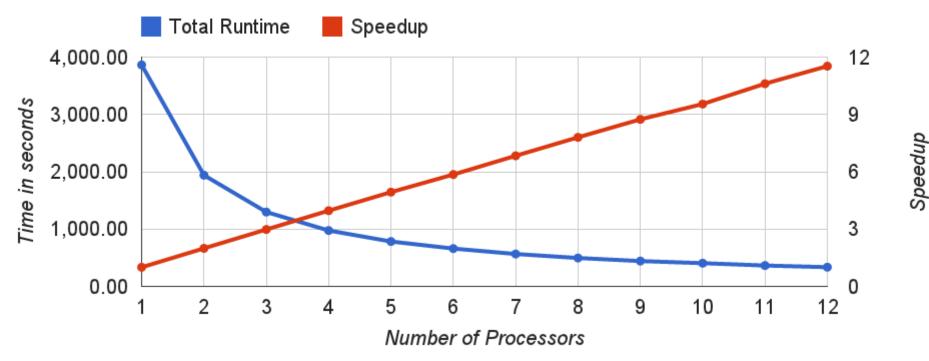


Speedup

• 335.18 seconds to find collision

Second Test: OpenMP on 12 Core System

- 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{32}$

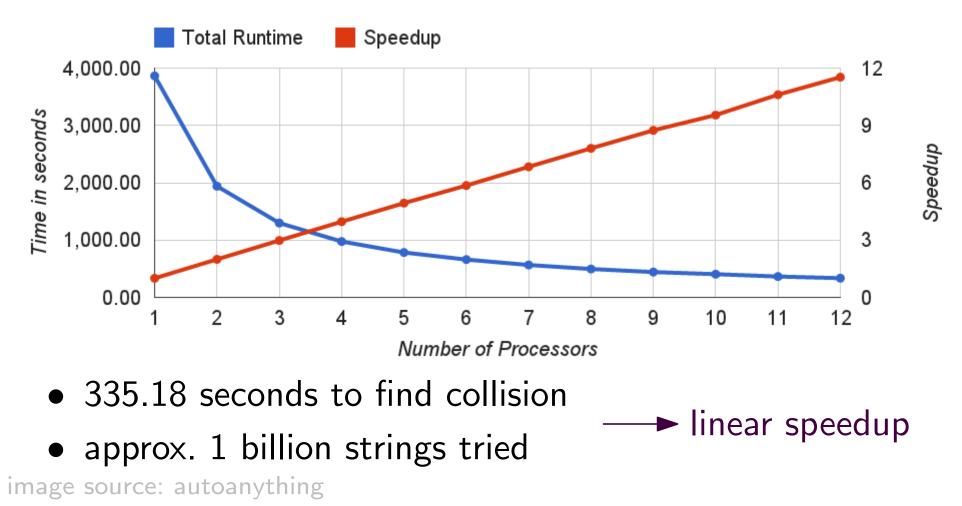


- 335.18 seconds to find collision
- approx. 1 billion strings tried



Second Test: OpenMP on 12 Core System

- 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{32}$







Third Test: OpenMP on 32 Core System









Third Test: OpenMP on 32 Core System

• 32 Intel Xeon E7-4830 at 2.13GHz

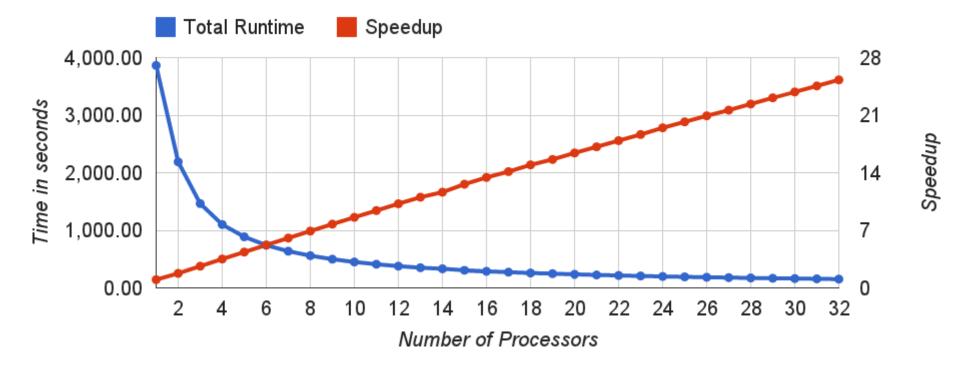
Third Test: OpenMP on 32 Core System



- 32 Intel Xeon E7-4830 at 2.13GHz
- input: md5(x) with $x \in \{0, 1\}^{32}$

Third Test: OpenMP on 32 Core System

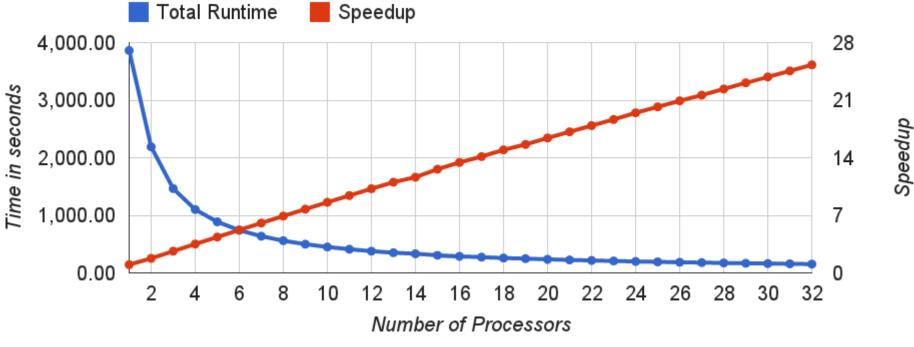
- 32 Intel Xeon E7-4830 at 2.13GHz
- input: md5(x) with $x \in \{0, 1\}^{32}$





Third Test: OpenMP on 32 Core System

- 32 Intel Xeon E7-4830 at 2.13GHz
- input: md5(x) with $x \in \{0, 1\}^{32}$

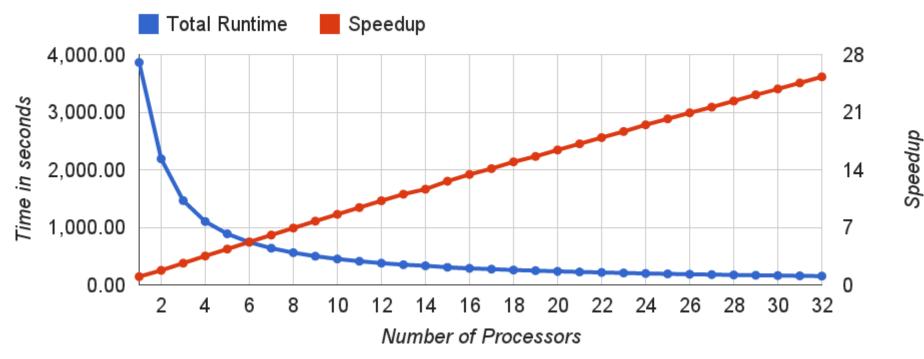


• 152.74 seconds to find collision



Third Test: OpenMP on 32 Core System

- 32 Intel Xeon E7-4830 at 2.13GHz
- input: md5(x) with $x \in \{0,1\}^{32}$

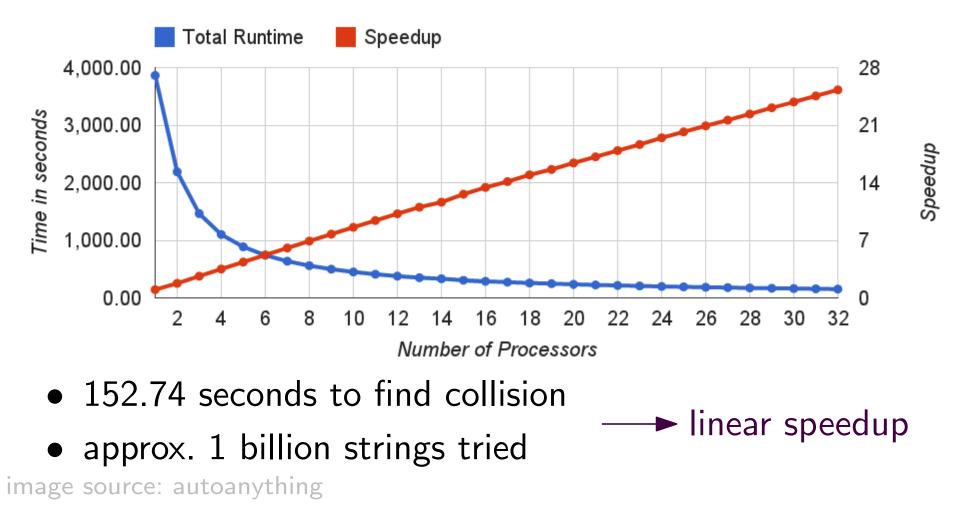


- 152.74 seconds to find collision
- approx. 1 billion strings tried



Third Test: OpenMP on 32 Core System

- 32 Intel Xeon E7-4830 at 2.13GHz
- input: md5(x) with $x \in \{0,1\}^{32}$









Fourth Test: MPI on 6 · 12 Core Systems







Benchmarks Fourth Test: MPI on 6 · 12 Core Systems



• 6 hosts with 12 Intel Xeon E5645 at 2.40GHz

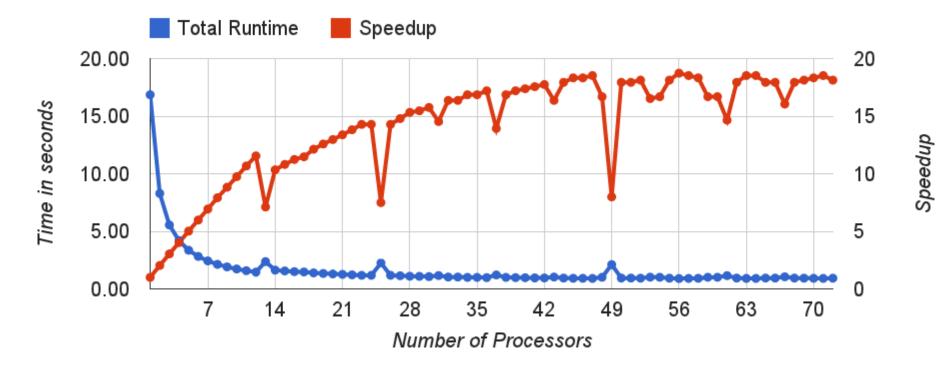
Fourth Test: MPI on 6 · 12 Core Systems



- 6 hosts with 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0,1\}^{24}$

Fourth Test: MPI on 6 · 12 Core Systems

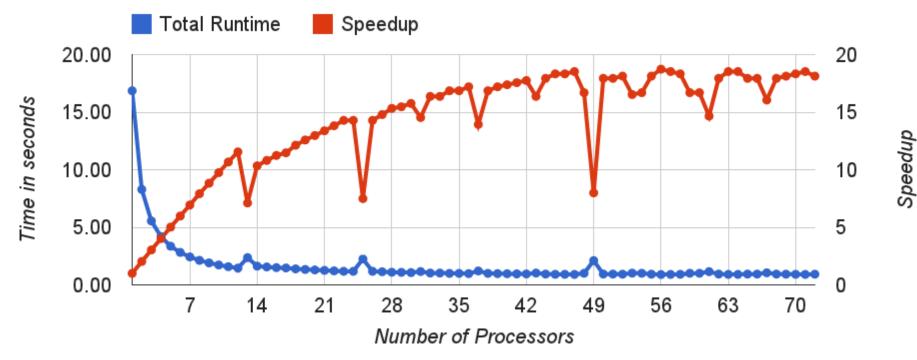
- 6 hosts with 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{24}$





Fourth Test: MPI on 6 · 12 Core Systems

- 6 hosts with 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{24}$

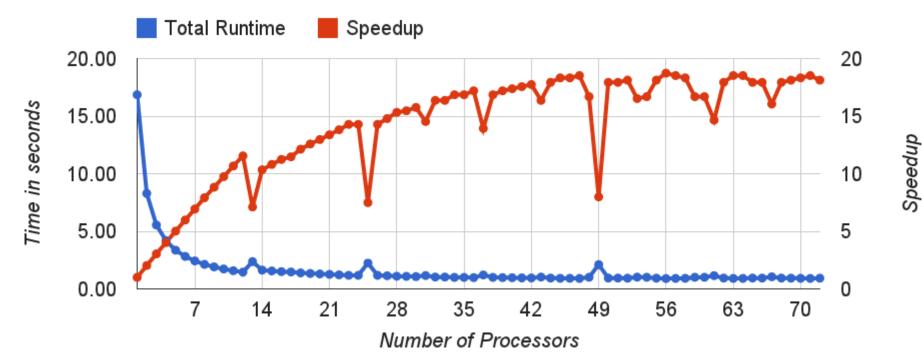


• 0.93 seconds to find collision



Fourth Test: MPI on 6 · 12 Core Systems

- 6 hosts with 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{24}$

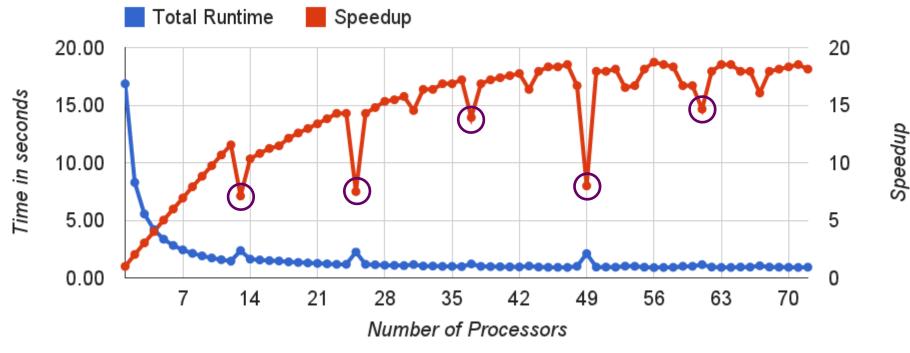


- 0.93 seconds to find collision
- approx. 4 million strings tried



Fourth Test: MPI on 6 · 12 Core Systems

- 6 hosts with 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{24}$

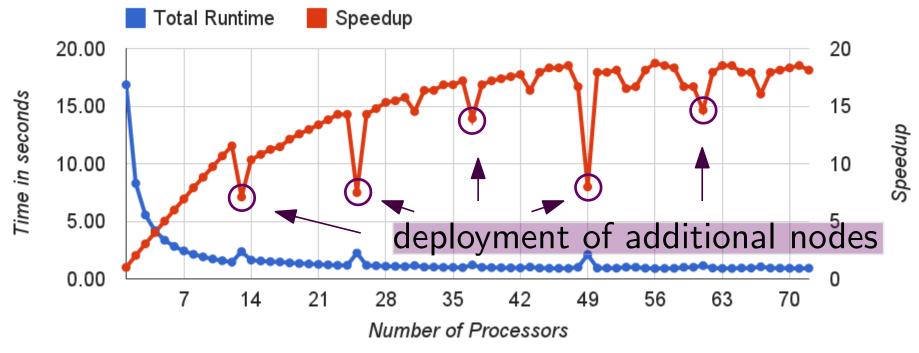


- 0.93 seconds to find collision —> speed drop
- approx. 4 million strings tried



Fourth Test: MPI on 6 · 12 Core Systems

- 6 hosts with 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0,1\}^{24}$

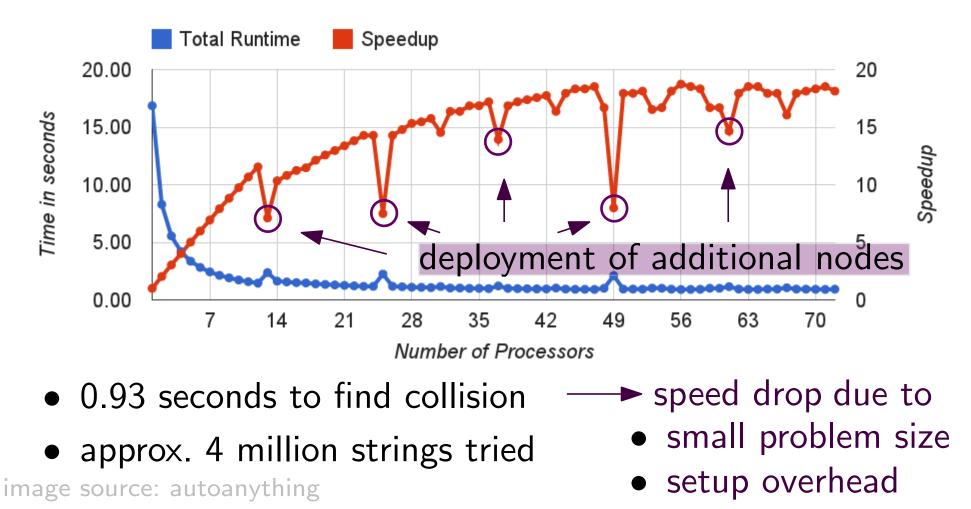


- 0.93 seconds to find collision —> speed drop
- approx. 4 million strings tried



Fourth Test: MPI on 6 · 12 Core Systems

- 6 hosts with 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{24}$









Fifth Test: MPI on 6 · 12 Core Systems



Fifth Test: MPI on 6 · 12 Core Systems

• 6 hosts with 12 Intel Xeon E5645 at 2.40GHz



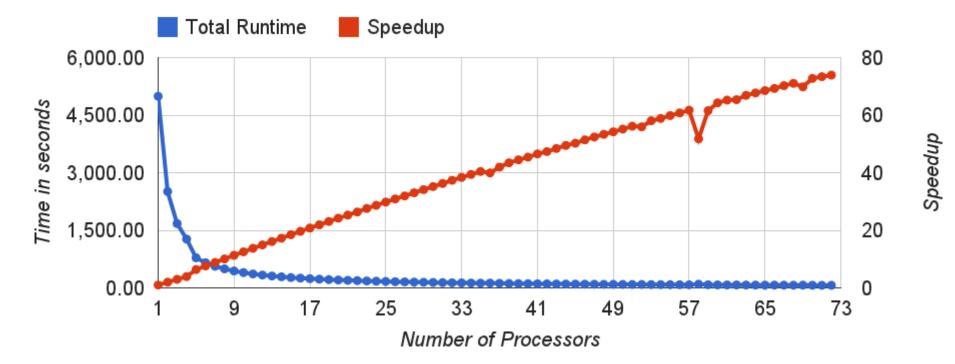
Fifth Test: MPI on 6 · 12 Core Systems



- 6 hosts with 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{32}$

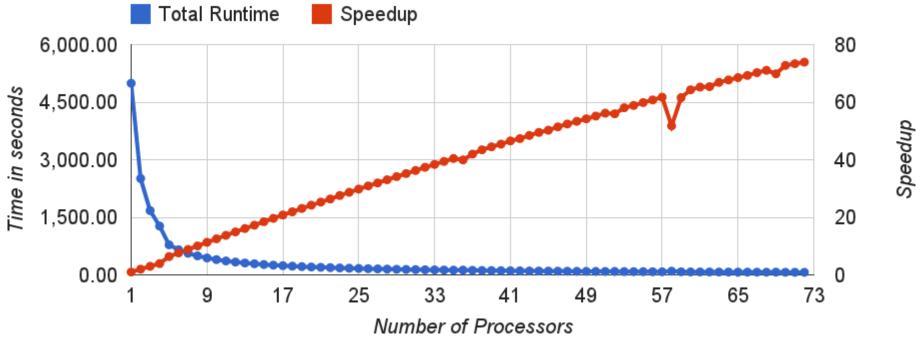
Fifth Test: MPI on 6 · 12 Core Systems

- 6 hosts with 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{32}$



Fifth Test: MPI on 6 · 12 Core Systems

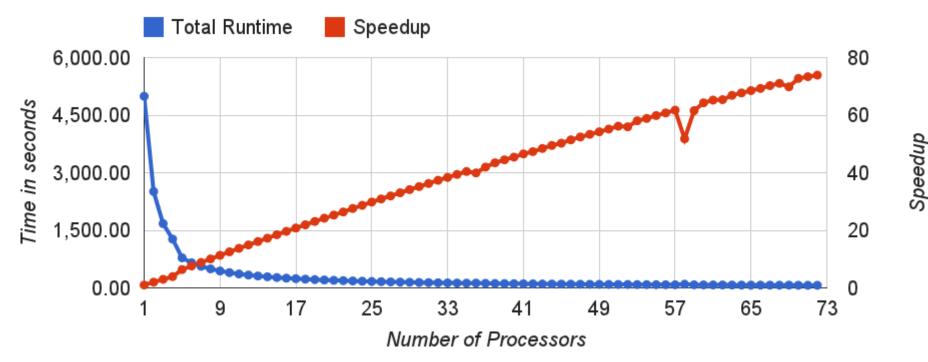
- 6 hosts with 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{32}$



• 67.54 seconds to find collision

Fifth Test: MPI on 6 · 12 Core Systems

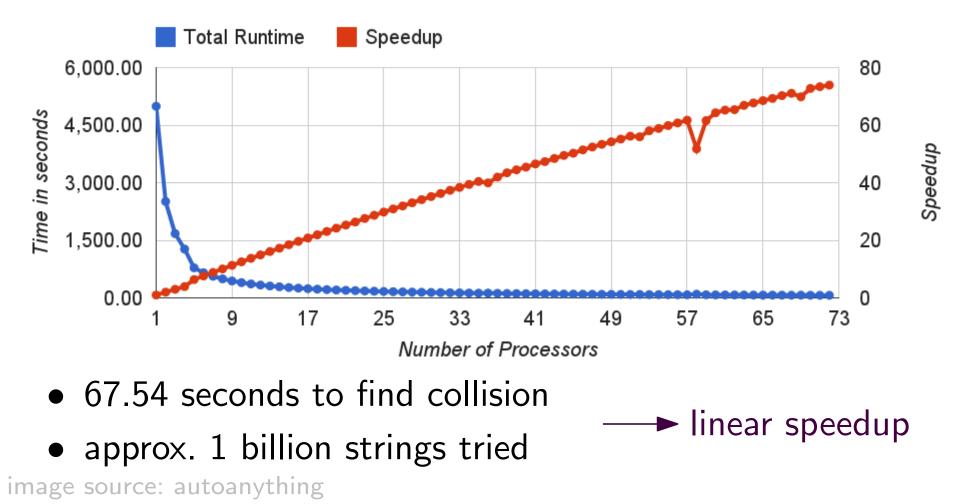
- 6 hosts with 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{32}$



- 67.54 seconds to find collision
- approx. 1 billion strings tried

Fifth Test: MPI on 6 · 12 Core Systems

- 6 hosts with 12 Intel Xeon E5645 at 2.40GHz
- input: md5(x) with $x \in \{0, 1\}^{32}$







Sixth Test: MPI on 2 · 32 Core Systems



Benchmarks Sixth Test: MPI on 2 · (32) Core Systems



Sixth Test: MPI on 2 · (32) Core Systems



• 2 hosts with 32 Xeon E7-4830 at 2.13GHz

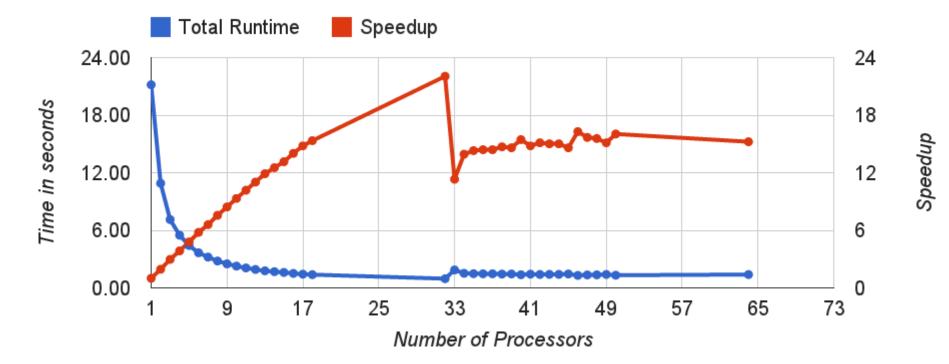
Sixth Test: MPI on 2 · (32) Core Systems



- 2 hosts with 32 Xeon E7-4830 at 2.13GHz
- input: md5(x) with $x \in \{0, 1\}^{24}$

Sixth Test: MPI on 2 · 32 Core Systems

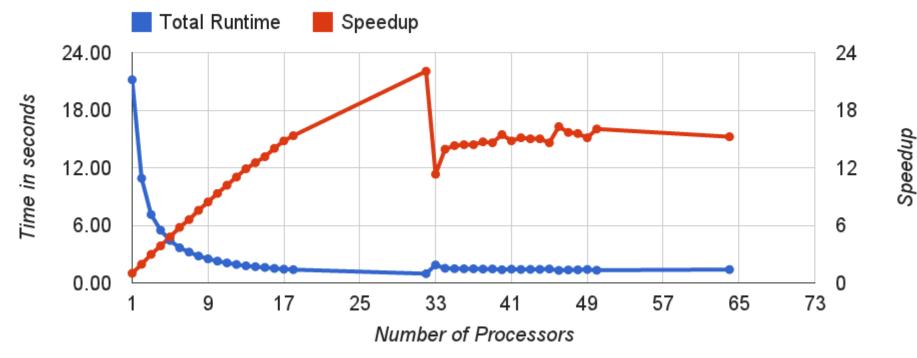
- 2 hosts with 32 Xeon E7-4830 at 2.13GHz
- input: md5(x) with $x \in \{0,1\}^{24}$





Sixth Test: MPI on 2 · 32 Core Systems

- 2 hosts with 32 Xeon E7-4830 at 2.13GHz
- input: md5(x) with $x \in \{0, 1\}^{24}$

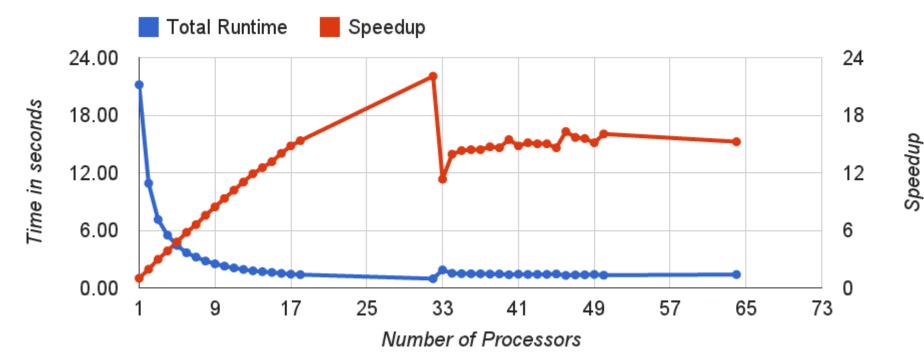


• 0.96 seconds to find collision



Sixth Test: MPI on 2 · 32 Core Systems

- 2 hosts with 32 Xeon E7-4830 at 2.13GHz
- input: md5(x) with $x \in \{0,1\}^{24}$

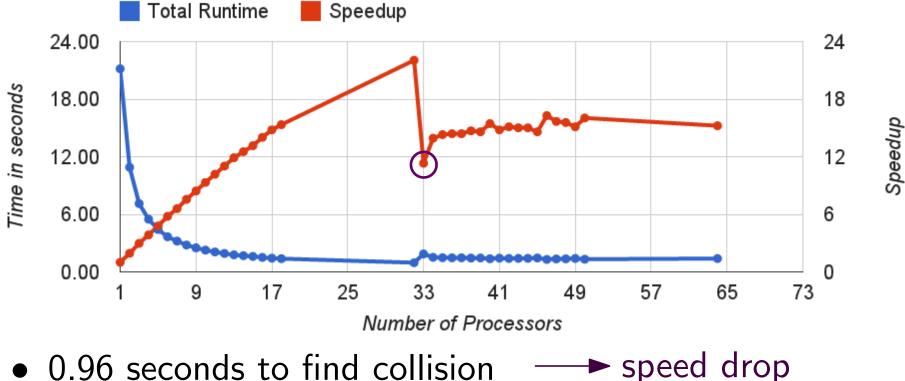


- 0.96 seconds to find collision
- approx. 4 million strings tried



Sixth Test: MPI on 2 · 32 Core Systems

- 2 hosts with 32 Xeon E7-4830 at 2.13GHz
- input: md5(x) with $x \in \{0, 1\}^{24}$

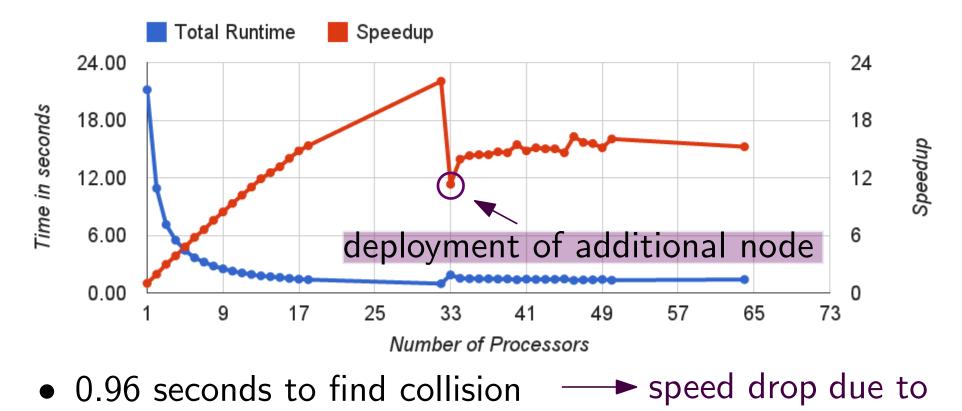


- 0.96 seconds to find collision
- approx. 4 million strings tried



Sixth Test: MPI on 2 · 32 Core Systems

- 2 hosts with 32 Xeon E7-4830 at 2.13GHz
- input: md5(x) with $x \in \{0,1\}^{24}$

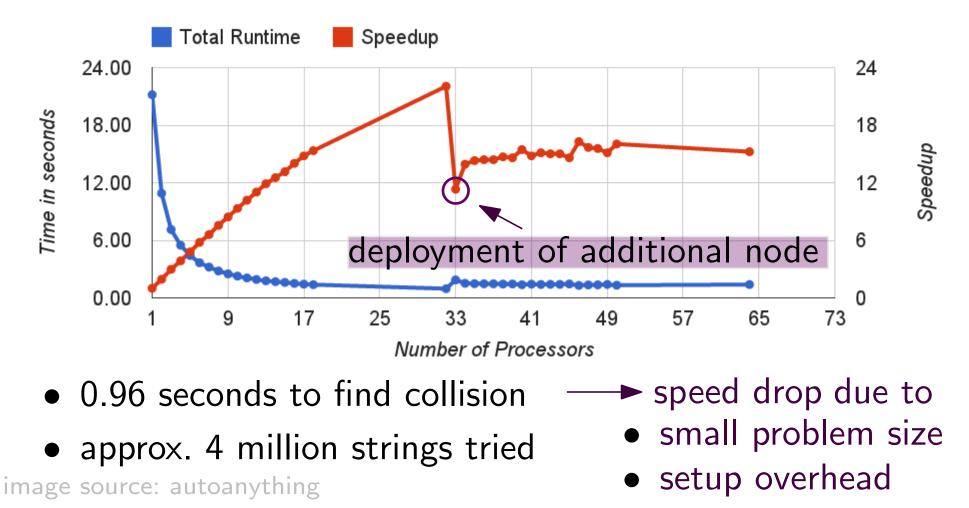


• approx. 4 million strings tried



Sixth Test: MPI on 2 · 32 Core Systems

- 2 hosts with 32 Xeon E7-4830 at 2.13GHz
- input: md5(x) with $x \in \{0,1\}^{24}$







Seventh Test: MPI on 2 · 32 Core Systems

FUEL

FUFL Seventh Test: MPI on 2 · 32 Core Systems

• 2 hosts with 32 Xeon E7-4830 at 2.13GHz

Seventh Test: MPI on 2 · 32 Core Systems

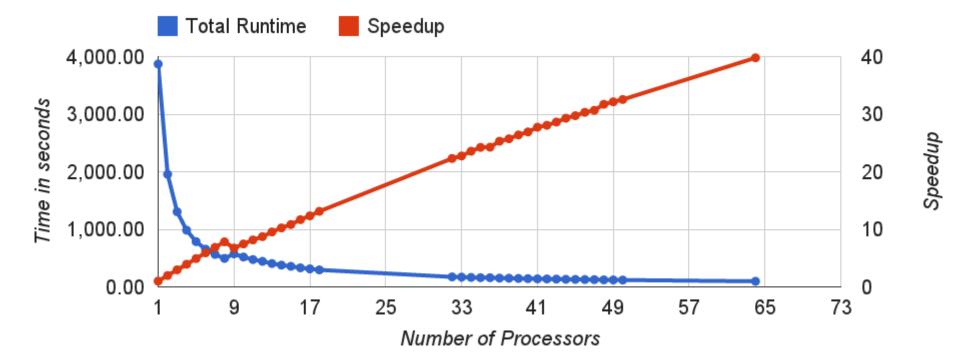
- 2 hosts with 32 Xeon E7-4830 at 2.13GHz
- input: md5(x) with $x \in \{0, 1\}^{32}$



ems

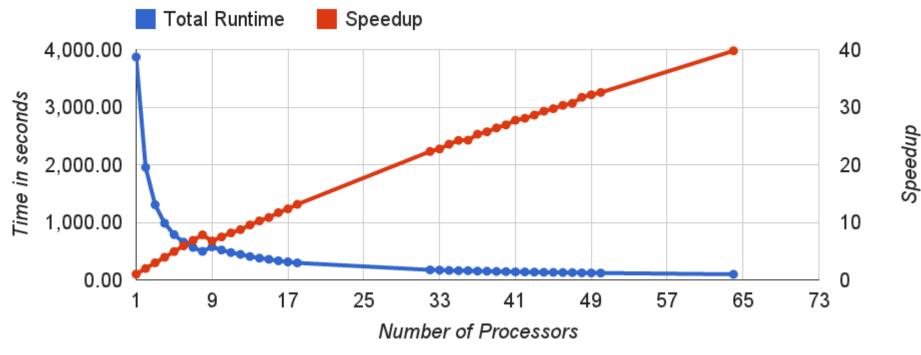
Seventh Test: MPI on 2 · 32 Core Systems

- 2 hosts with 32 Xeon E7-4830 at 2.13GHz
- input: md5(x) with $x \in \{0, 1\}^{32}$



Seventh Test: MPI on 2 · 32 Core Systems

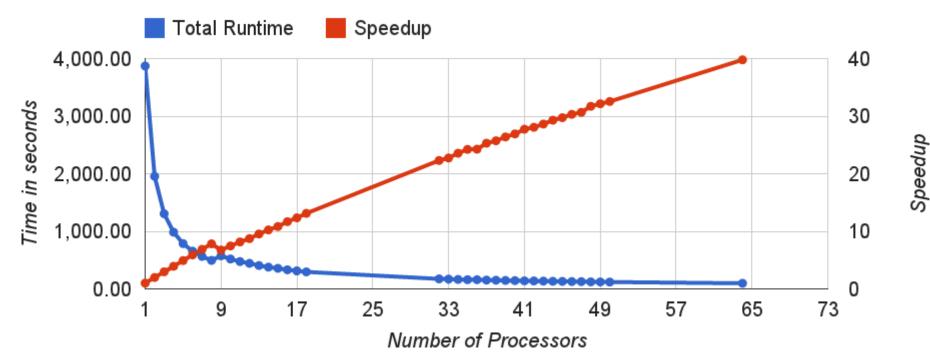
- 2 hosts with 32 Xeon E7-4830 at 2.13GHz
- input: md5(x) with $x \in \{0, 1\}^{32}$



• 97.30 seconds to find collision

Seventh Test: MPI on 2 · 32 Core Systems

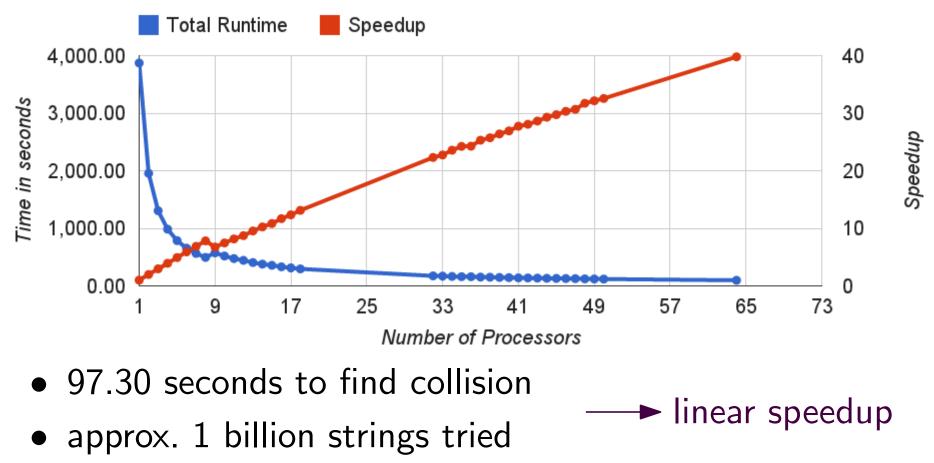
- 2 hosts with 32 Xeon E7-4830 at 2.13GHz
- input: md5(x) with $x \in \{0, 1\}^{32}$



- 97.30 seconds to find collision
- approx. 1 billion strings tried

Seventh Test: MPI on 2 · 32 Core Systems

- 2 hosts with 32 Xeon E7-4830 at 2.13GHz
- input: md5(x) with $x \in \{0, 1\}^{32}$





OpenMP implementation



OpenMP implementation

• linear speedup due to few communication



OpenMP implementation

- linear speedup due to few communication
- suitable for smaller problem sizes



OpenMP implementation

- linear speedup due to few communication
- suitable for smaller problem sizes
 but: already used CCR's "biggest" machine



OpenMP implementation

- linear speedup due to few communication
- suitable for smaller problem sizes but: already used CCR's "biggest" machine

MPI implementation



OpenMP implementation

- linear speedup due to few communication
- suitable for smaller problem sizes
 but: already used CCR's "biggest" machine

MPI implementation

• slower communication, setup overhead



OpenMP implementation

- linear speedup due to few communication
- suitable for smaller problem sizes
 but: already used CCR's "biggest" machine

MPI implementation

- slower communication, setup overhead
- more processors available



OpenMP implementation

- linear speedup due to few communication
- suitable for smaller problem sizes
 but: already used CCR's "biggest" machine

MPI implementation

- slower communication, setup overhead
- more processors available

suitable for larger problem sizes



OpenMP implementation

- linear speedup due to few communication
- suitable for smaller problem sizes
 but: already used CCR's "biggest" machine

MPI implementation

image source: ozbinoculars

- slower communication, setup overhead
- more processors available

suitable for larger problem sizes



compare

OpenMP implementation

- linear speedup due to few communication
- suitable for smaller problem sizes
 but: already used CCR's "biggest" machine

MPI implementation

- slower communication, setup overhead
- more processors available

suitable for larger problem sizes



Compare $x \in \{0, 1\}^{32}$

OpenMP implementation

- linear speedup due to few communication
- suitable for smaller problem sizes
 but: already used CCR's "biggest" machine

MPI implementation

- slower communication, setup overhead
- more processors available

suitable for larger problem sizes



ad	compare $x \in \{0, 1\}^{32}$	
	OpenMP	
	MPI	

OpenMP implementation

- linear speedup due to few communication
- suitable for smaller problem sizes
 but: already used CCR's "biggest" machine

MPI implementation

- slower communication, setup overhead
- more processors available

suitable for larger problem sizes

compare $x \in \{0, 1\}^{32}$			
# PEs	32		
OpenMP			
MPI			



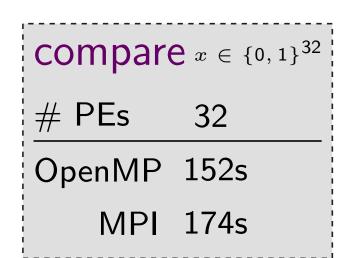
OpenMP implementation

- linear speedup due to few communication
- suitable for smaller problem sizes
 but: already used CCR's "biggest" machine

MPI implementation

- slower communication, setup overhead
- more processors available

suitable for larger problem sizes





OpenMP implementation

- linear speedup due to few communication
- suitable for smaller problem sizes
 but: already used CCR's "biggest" machine

MPI implementation

- slower communication, setup overhead
- more processors available

suitable for larger problem sizes





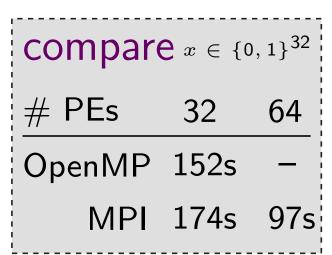
OpenMP implementation

- linear speedup due to few communication
- suitable for smaller problem sizes
 but: already used CCR's "biggest" machine

MPI implementation

- slower communication, setup overhead
- more processors available

suitable for larger problem sizes





OpenMP implementation

- linear speedup due to few communication
- suitable for smaller problem sizes
 but: already used CCR's "biggest" machine

MPI implementation

- slower communication, setup overhead
- more processors available

suitable for larger problem sizes

CUDA implementation



OpenMP implementation

- linear speedup due to few communication
- suitable for smaller problem sizes
 but: already used CCR's "biggest" machine

MPI implementation

- slower communication, setup overhead
- more processors available

suitable for larger problem sizes

CUDA implementation (future work)



OpenMP implementation

- linear speedup due to few communication
- suitable for smaller problem sizes
 but: already used CCR's "biggest" machine

MPI implementation

- slower communication, setup overhead
- more processors available

suitable for larger problem sizes

CUDA implementation (future work)

• fast communication between processing elements



OpenMP implementation

- linear speedup due to few communication
- suitable for smaller problem sizes
 but: already used CCR's "biggest" machine

MPI implementation

- slower communication, setup overhead
- more processors available

suitable for larger problem sizes

CUDA implementation (future work)

- fast communication between processing elements
- very high number of processors on single nodes





Hans Delf and Helmut Knebl. Introduction to Cryptography: Principles and Applications. Springer, 2007.

Alfred J. Menezes, Paul C. van Oorschot, Scott A. Vanstone. *Handbook of Applied Cryptography*. CRC Press, 1996.

Charles P. Pfleeger and Shari L. Pfleeger. *Analyzing Computer Security: A Threat/Vulnerability/Countermeasure Approach* Prentice Hall, 2011.