Horst-Görtz Institut
für IT Sicherheit

Efficient Hash Collision Search Strategies on Special-Purpose Hardware

Tim Güneysu, Christof Paar, Sven Schäge

Overview

- Properties of MD4-family hash functions
- Attacks on MD4-family hash functions
- Implementation requirements for collision search algorithms
- Implementation details
- Performance results
- Estimations for SHA-1
- Conclusion

Cryptographic Hashfunction

Function that efficiently maps arbitrarily long input to fixed-size output

Three properties

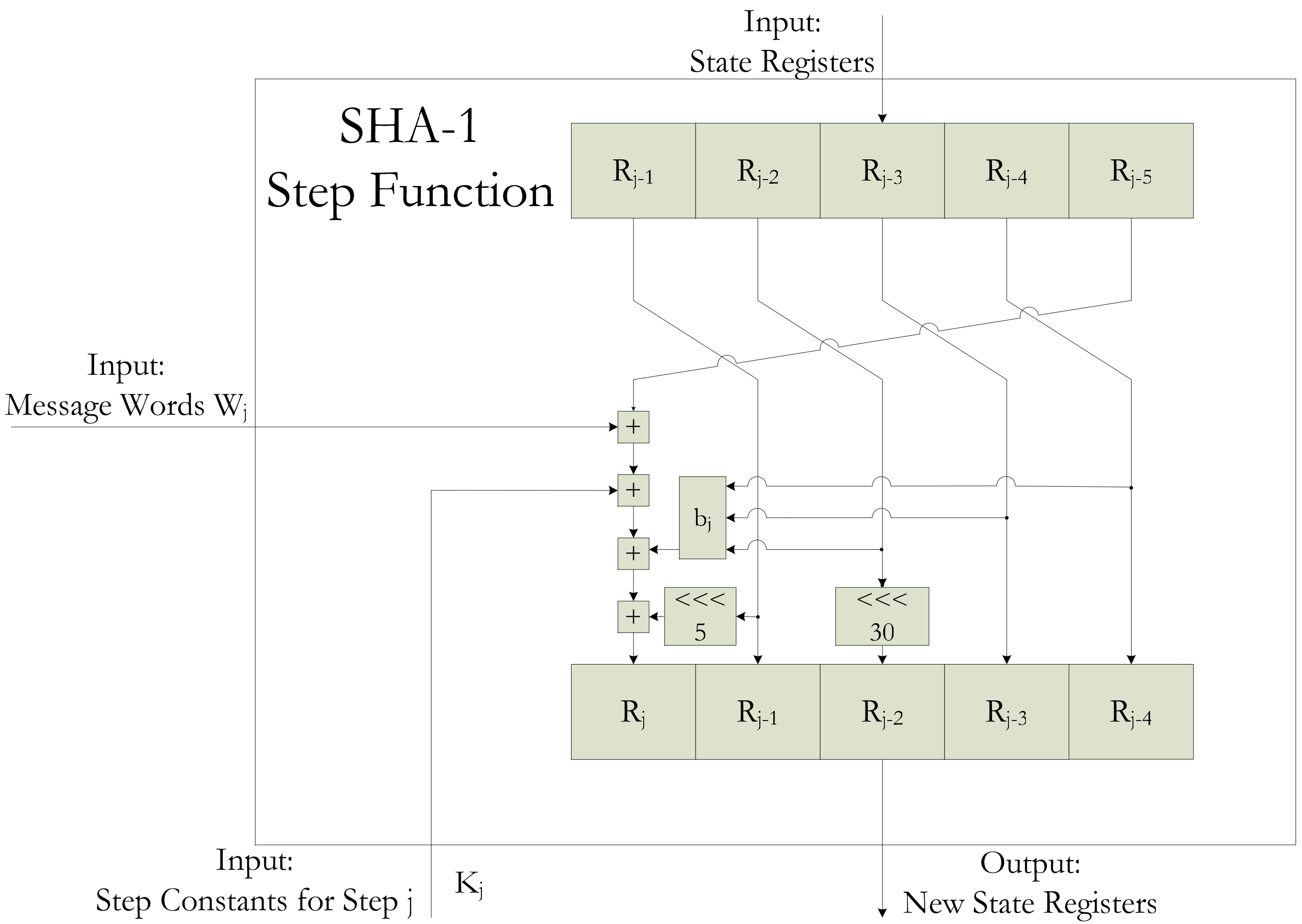
- Preimage resistance („one-way“), attack complexity: $O(2^n)$
- Second preimage resistance, attack complexity: $O(2^n)$
- Collision resistance, attack complexity: $O(2^{n/2})$

Practical problem: „arbitrarily long input“

Solution: Iteration

- Hash function makes one compression function invocation per message block
- Compression function iteratively applies step function to state variables (fixed iteration number)

Compression Function of SHA-1



Attacks on Compression Functions: Phase 1

Differential attack on the compression function

Collisions can adequately be described using differences

- Non-zero input and zero output difference

Search for conditions such that difference propagation can be controlled with sufficient probability

Attacks on Compression Function: Phase 2

Exploitation of remaining degree of freedom for concrete message choice

Use for partly predetermining message bits

Use for acceleration of collision search

- Find bits in the computation path that can (indirectly) be altered without disturbing previous, yet satisfied conditions
- From a collision candidate fulfilling all conditions so far a new candidate with the same characteristic can be computed
- Computing new candidates is much more efficient than randomly choosing new messages and testing for compliance with conditions
- Number of candidates grows exponentially with number of found bits

Collision Search Algorithms

Collision Algorithms for MD5

Algorithms that extensively applies acceleration techniques have best speed results

Fastest Algorithm (CS) [Klima06]:

- Complexity of about 2^{33} MD5 step operations [Joŝĉák06]
- Pentium 4, 2 GHz, 30 s on average to find a collision

Hints on Implementation Requirements

Probabilistic search

- Pseudo random number generator (PRNG)

MD4-family hash functions developed for 32-bit (64-bit) processors

- Collision search algorithms should also work on 32-bit (64-bit) units
- Otherwise, expensive correction operations are required

Most operations on lower hierarchical levels process the result of their immediate predecessor

- Parallelization on lower hierarchical levels is hardly useful

Acceleration techniques are very effective

- Require additional branches and loops in the computation path
- Usual hardware acceleration techniques like pipelining not useful

Implementation

µMD: 32-bit ASIC microprocessor

- General-purpose for MD4-family hash functions
- Just 16 native instructions

µCS: full collision search unit based on µMD

- Equipped with necessary memory and I/O logic

Dedicated assembler for µMD

Implementation of CS for µCS ROM

- Performance tests

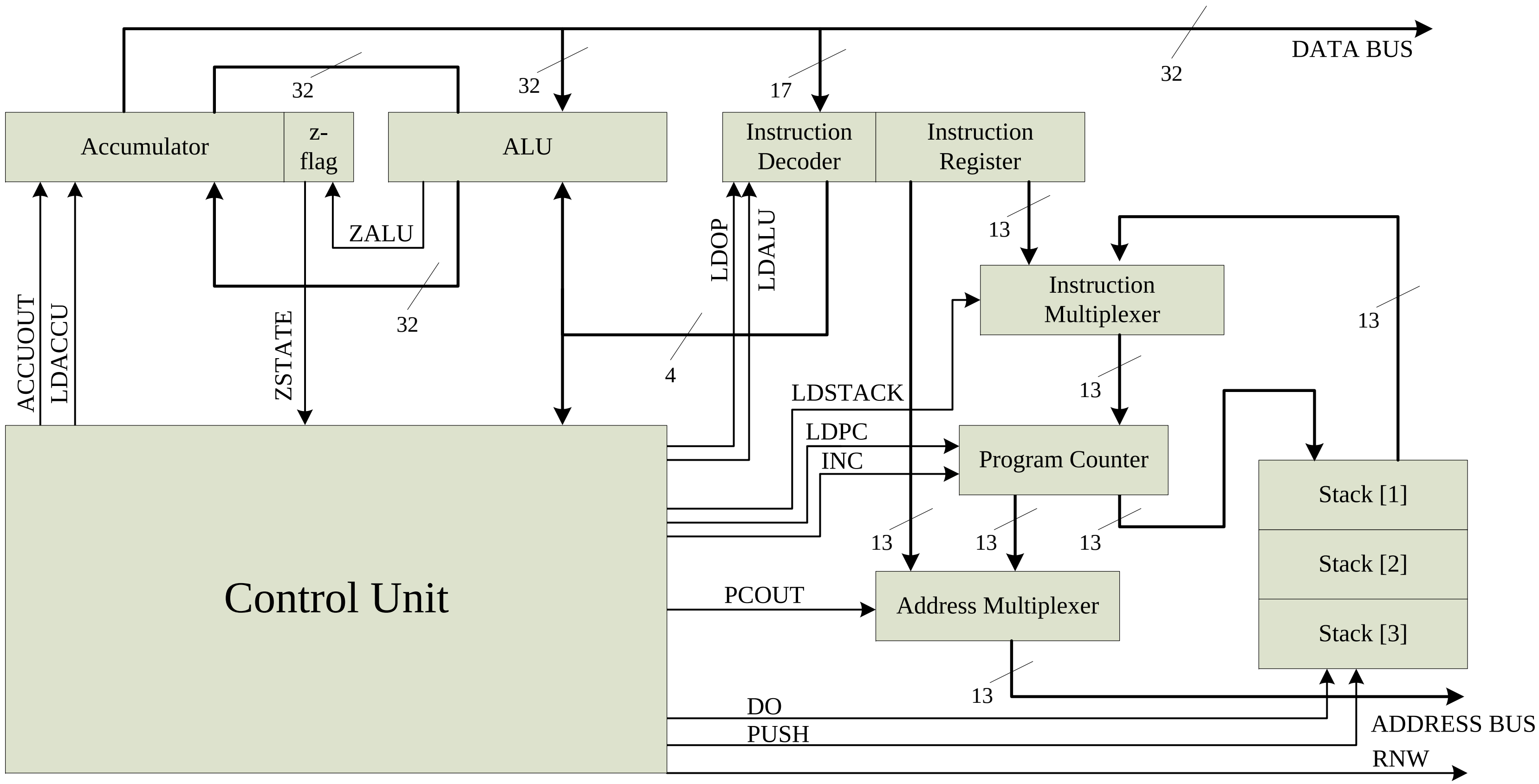
Instruction Set

Most operations require 2 cycles for execution

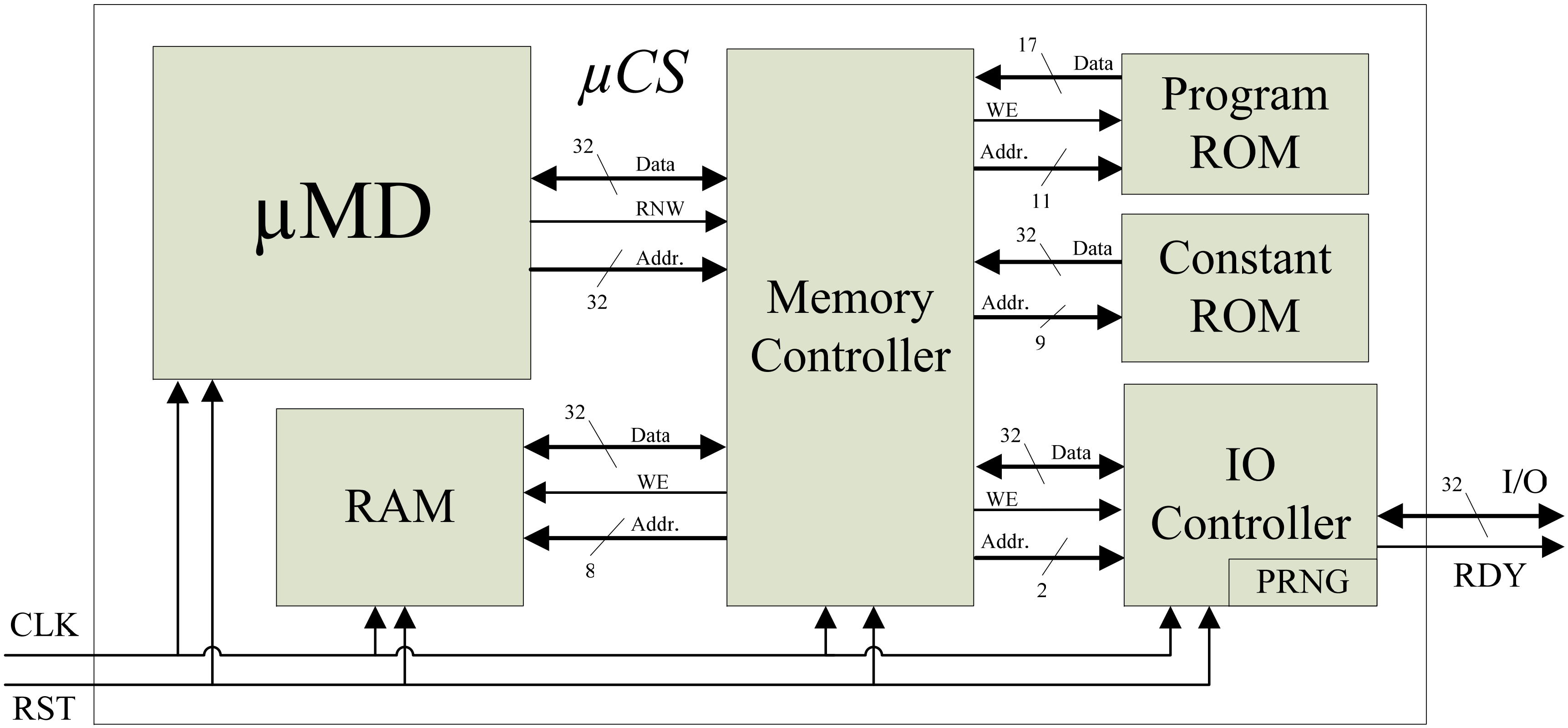
LDI, STI, CALL, RET are used for parameterized subfunction calls

STA	STI
LDA	LDI
ADD	CALL
SUB	RET
NOT	JMP
AND	JE
OR	JNE
XOR	RL

µMD:Overview



µCS: Overview



Implementation Results

	Pentium 4, 2 GHz	μMD	μCS
clock frequency	2 GHz	228.8 MHz	102.9 MHz
cycles to find a collision	$60*10^9$	$480*10^9$	$480*10^9$
time to find a collision	30 s	2097.6 s	4660.8 s
area	146 mm^2	0.027 mm^2	0.960 mm^2
area-time product	4380	55.9	4472.6

Parallelization

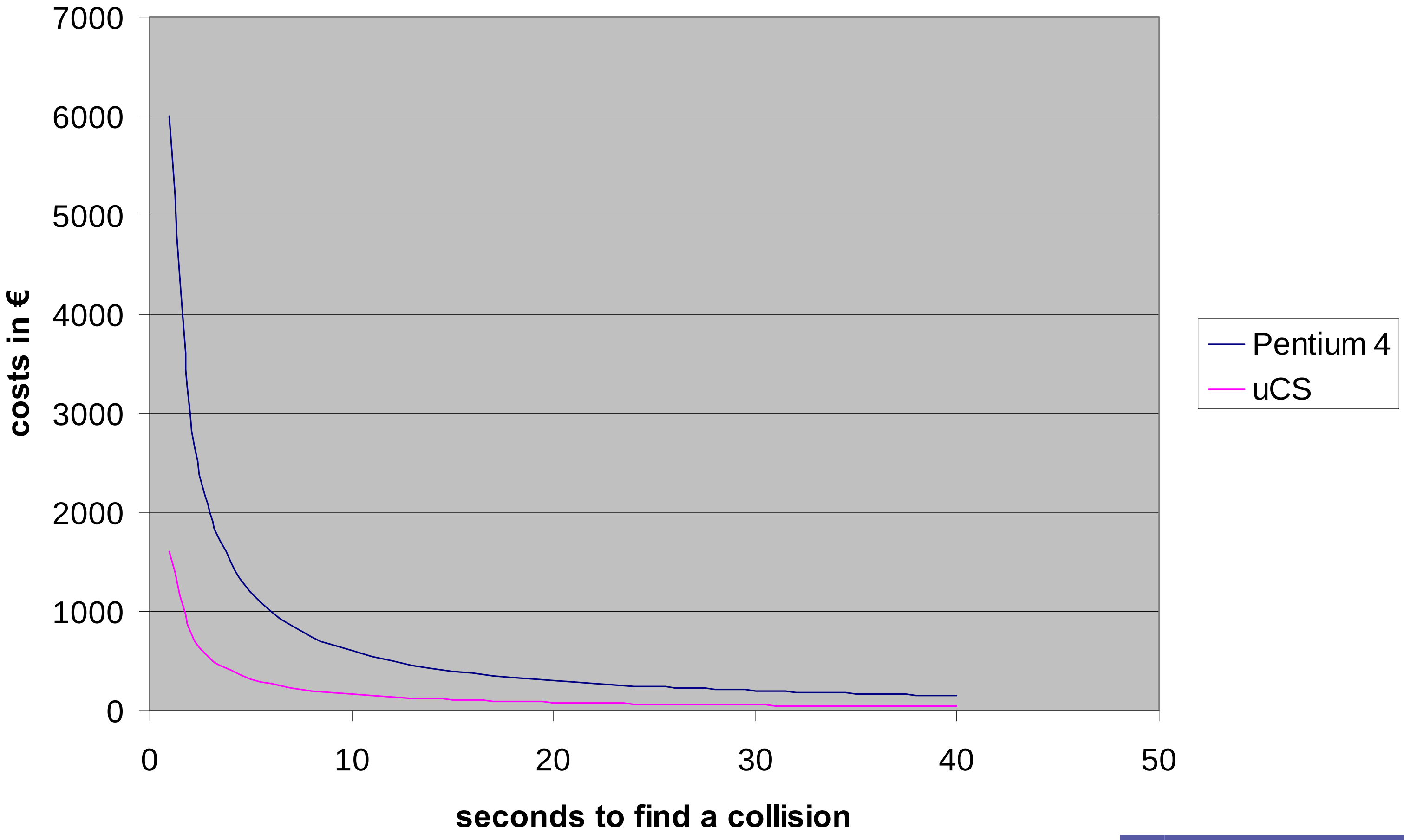
Assumptions

- Costs for Pentium 4, 2 GHz processor: 50 €
- Costs for off-the-shelf parallelization equipment: 150 €, 300 % overhead
  - Motherboard (PXE), network card: 80 €, fan: 12 €, RAM: 25 €, power supply: 25 €, additional equipment for network infrastructure
- Overall costs for a parallelized Pentium 4 system: 200 €
- Pentium 4 and µCS have equal production constraints
  - Same price per circuit area 50 €/146 mm^2 = 0,3425 €/mm^2
- Costs for parallelizing µCS units is negligible
  - Less than 5 %
- Overall costs for a parallelized µCS unit: 0.35 €

After parallelization, finding a MD5 collision in 1 s costs…

- 6000 € when invested in a Pentium 4 based architecture
- 1608.4 € when invested in a µCS based architecture

Comparison between µCS and Pentium 4



Estimations for SHA-1

Current SHA-1 attack complexity: about 2^{62} compression function evaluations [Wang06]

- Roughly 2^{70} step function evaluations

Assumption

- MD5 and SHA-1 step operations have equal execution time

SHA-1 attack is 2^{37} times slower than MD5 attack

Given 1 mio. €, finding a SHA-1 collision would take…

- 26 years when invested in Pentium 4 architecture
- 7 years when invested in µCS architecture

Conclusion

Horst-Görtz Institut für IT Sicherheit

Using special-purpose hardware for collision search pays off

µCS is roughly 3.7 times better than Pentium 4 based architectures

There is much space for improvements

- Advanced processor features can be added to our design
- New features can easily be evaluated concerning their efficiency gain for collision search

End

Thank you for your attention.

Any questions?