



QUANTIFYING THE STRENGTH OF HASH FUNCTIONS



By
Alexander Adelholm Brandbyge
&
Lars Embøll Nielsen

Supervisor: *Christian Damsgaard Jensen*

PAGE 1 OF 134
SIGNATURES
January 11th 2016

ALEXANDER ADELHOLM BRANDBYGE

LARS EMBØLL NIELSEN

Technical University of Denmark
DTU Compute, Building 324
Richard Petersens Plads
2800 Kongens Lyngby
&
Technical University of Denmark
DTU Transport, Building 115
Bygningstorvet
2800 Kongens Lyngby

Kongens Lyngby 11/1-2016



FOREWORD

To all whom it may concern:

Be it known that we, ALEXANDER ADELHOLM BRANDBYGE & LARS EMBØLL NIELSEN, subjects of Her Majesty Margrethe the Second, queen regnant of Denmark, Greenland and the Danish Dominions beyond the Seas, from Vejen , Jutland & Sorø, Sealand, border country of Germany, residing at the royal city of Lyngby, in the Capital Region of Denmark, have invented certain new and useful Improvements in QUANTIFYING THE STRENGTH OF HASH FUNCTIONS, of which the following is a specification, reference being had to the drawings accompanying and forming a part of the same.

TABLE OF CONTENTS

0 Introduction	8
0.1 What is a hash?.....	9
0.2 Report Structure	10
1 Common Cryptographic Hash Algorithms	11
1.1 MD5.....	11
1.2 SHA-1	12
1.3 SHA-2	12
1.4 SHA-3	12
2 Risk Management Theory.....	13
2.1 Bathtub & Rocking Boat	14
2.1.1 Rocking Boat Principle	15
2.2 Risk Statistics	15
2.2.1 Near Misses & Bug Reports	17
2.2.1.1 The Incident Pyramid	17
2.2.1.2 Responsible Disclosure in a Risk Assessment Perspective	18
2.3 Risk Acceptance	19
2.3.1 Railway Safety	21
2.3.2 SIL & IEC 61508.....	21
2.3.3 European Railway Security	22
2.4 Risk Classification	23
2.4.1 The money value of a man	23
2.4.1.1 Human Capital (HK).....	23
2.4.1.2 Value of a Statistical Life (VSL)	24
2.4.1.3 Value of Preventing a Casualty	25
2.4.1.4 ALARP	26
2.4.2 Epistemic uncertainty in Danish VSL	27
3 Collision Probability	27



- 3.1 Hash Collisions.....27
- 3.2 Brute Force28
- 3.3 Applied Birthday Paradox29
- 3.4 Moore’s Law.....31
- 4 X.509 Structure32
 - 4.1 Encapsulation32
 - 4.2 Certificate modification34
- 5 Current use of SHA-134
 - 5.1 Code Signing35
 - 5.2 Document Signing35
 - 5.3 BitTorrent Protocol35
 - 5.3.1 BitTorrent Metadata Files36
 - 5.3.2 Structure36
 - 5.3.2.1 Info.....37
 - 5.3.2.2 Typical BitTorrent File.....37
 - 5.3.3 Tracker protocol37
 - 5.3.4 Peer Protocol39
 - 5.3.5 DHT.....39
 - 5.3.5.1 Usage in BitTorrenting.....39
 - 5.3.6 PEX40
 - 5.4 Content Distribution Networks40
 - 5.5 openPGP40
 - 5.6 Law41
 - 5.7 Summary of KPI42
- 6 Hazard Identification43
 - 6.1 Apple Update Distribution.....43
 - 6.2 Document Signing43
 - 6.3 Certificates.....44
 - 6.3.1 Trust 2408.....45
 - 6.4 BitTorrent46
 - 6.4.1 Fake-block Attack.....46
 - 6.4.2 Uncooperative-peer Attack46
 - 6.4.3 Leeching47
 - 6.4.4 Torrent Availability.....47
 - 6.5 Peer to Peer47
 - 6.6 End to End.....48
- 7 GPU SHA-1 Collision Probability Estimate49
 - 7.1 Design considerations.....49



- 7.1.1 HPC Forcer Architecture49
- 7.1.2 SHA-1 Kernel50
- 7.1.3 Optimizations52
- 7.1.4 Prehash Value52
- 7.2 GPGPU & CUDA52
 - 7.2.1 Language Variations53
 - 7.2.2 Available Hardware53
 - 7.2.3 Core concepts.....54
 - 7.2.4 Memory model.....57
 - 7.2.5 CUDA C/C++ specifics58
- 8 BitTorrent60
 - 8.1 Data source.....60
 - 8.2 BitSnoop Data extraction60
 - 8.3 Magnet link resolution.....61
- 9 When will we see a SHA-1 collision?61
- 10 SHA-1 Collision Testing.....62
 - 10.1 HPC Diagnostics62
 - 10.2 HPC SHA-1 generation62
 - 10.3 HPC Evaluation.....63
 - 10.4 Applied Pigeonhole.....64
 - 10.5 BitTorrent SHA-1 Extraction.....65
 - 10.6 Evaluation of the BitTorrent SHA-1 Source.....68
- 11 Alternative Attack Vectors69
 - 11.1 Railway Methodologies69
 - 11.1.1 Safe Link Layer69
 - 11.1.2 Low Entropy Session Identification71
 - 11.1.3 American Railway Risk Model72
 - 11.1.4 Open ETCS.....72
 - 11.2 NemID73
 - 11.2.1 SHA-1 Root Certificate Verification75
- 12 Impact Analysis77
 - 12.1 Denial of Service78
 - 12.2 Railway80
 - 12.3 Heartbleed84
 - 12.4 Chapter Summary86
- 13 Consequence Analysis.....87
 - 13.1 Random Data Collision Within One Hour87
 - 13.2 Specific Data Collision Within One Hour88



- 14 Risk Evaluation88
 - 14.1 Schneier misunderstanding Stevens90
 - 14.2 Analysis on the estimates derived from own data90
- 15 Risk mitigation: Responsible Disclosure91
 - 15.1 Storing Secrets Securely95
 - 15.1.1 Shamir Secret Sharing95
 - 15.1.2 Setup95
 - 15.1.3 Other usage96
- 16 Summary of Part 396
- 17 Conclusion97
 - 17.1 Recommendations98
 - 17.1.1 Future projects should use SHA-398
 - 17.1.2 Authentication Message Entropy98
 - 17.1.3 OpenPGP RFC 488099
 - 17.1.4 Certificate Transparency99
 - 17.1.5 Flexibility in security critical container types99
 - 17.1.6 Tip on Good Hash100
 - 17.2 Future Work100
 - 17.2.1 HPC100
 - 17.2.2 Torrent100
 - 17.2.3 Data on SHA-1 usage101
- 18 Bibliography102
- 19 Abbreviations, technical terms & definitions111
 - 19.1 Abbreviations111
 - 19.2 Technical terms and definitions112
 - 19.3 Units & numbers114
 - 19.3.1 Short number scale114
 - 19.3.2 Metric prefixes114
 - 19.3.3 Binary prefixes114
 - 19.3.4 SI units114
 - 19.3.4.1 Derived:114
- 20 Appendix115
 - 20.1 Example Certificate115
 - 20.1.1 Modified certificate overview115
 - 20.1.1.1 Original certificate119
 - 20.2 Bencoding123
 - 20.3 HPC Platform Deployment123
 - 20.3.1 Job scripts123



20.3.2 ABACUS Scripts	124
20.4 Shamir Secret Sharing Toolkit Readme	125
20.5 ERA letters	126
20.5.1 Letter 1, December 2 nd 12:02.....	126
20.5.2 Letter 2, December 3 rd 11:49	129



ABSTRACT

In an estimate made by Bruce Schneier, it is predicted that the SHA-1 Hash algorithm will be cryptographically broken within the year 2018. This has will have a huge impact on the security infrastructure used today as SHA-1 is used extensively in many areas.

The report will outline the major areas where SHA-1 is used and offer a risk analysis based on theoretical models, previous examples and a practical implementation on a high performance computing cluster, and while no concrete, working attacks were produced, the hardware capabilities of the current generation were demonstrated, and used to reinforce the point, that 2nd pre-image attacks on SHA-1 are still not possible.

Intended Audience

The intended audience for this report are those who have obtained at least a bachelor's degree in computer science bachelor or better, for that reasons terms and concepts like "string", "integer", public-key cryptography and attackvector are not described and are assumed to be known or understandable with a quick internet search.

ACKNOWLEDGEMENTS

-ALEXANDER

I would like to thank my family and friends who have supported me through my life and education, I would not be where I am today if I had not been given the encouragement and help you all have provided.

Special thanks goes out to David Johannes Christensen for our endless talks of both practical and theoretical security, and for keeping up with my ramblings whenever I needed someone to help me gather my thoughts. And finally, Loreta Bllaci for being there for me no matter what.

-LARS

Thanks goes out to:

Lars Schiøtt Sørensen – for the introduction to fire-safety and by that the economical evaluation methods in assessing the value of a statistical life.

Stefan Lindhard Mabit – for an introduction to Discrete Choice Modelling and the wonders of interpreting statistical data.

Ismir Mulalic – for giving insight to economics and a deeper, profound interest in Discrete Choice Modelling.

Igor Kozine – not only a great teacher sparking an interest in System Safety and Reliability Engineering, but also a great person.

Per Bruun Brockhoff – for a vivid introduction into statistics, and its application in everyday life.

Susanne Vennerstrøm - for a nice and challenging introduction in astrophysics.

Per Høeg - for the stories of ESA, Galileo and inspiration for new students in the field of global positioning.

Jørgen Bo Christensen - for helping in the human factor of dealing with studying as well as how to handle having the responsibility of other people's lives as an engineer.

Gunnar Bagge - for an introduction in soil mechanics and establishment of engineering to be a field for safety analytics.

Kurt Kielsgaard Hansen - for invoking a curiosity in the world around us, and showing an empirical approach the challenges presented.

Jens Eising & Carsten Thomassen– for teaching an understanding and love of math, rather than just formulas.

Gregory Bell - for a view into management from the perspective of US department of Energy / ESnet.

Kjeld Nielsen R.I.P. - for an introduction into Facilities Management and lifecycle costs.

Torben Holvad - for welcoming and encouraging a project involving ERA data.



Carl Sagan, Brian Cox & Richard Feynman – for being inspirations of how science is building the foundation of the future and hence need glorious goals for us to know what direction to build in. Also for showing science should not stay in basements, but be liberated and expressed truthfully, in a way humanity as a whole can understand. René Xavier Victor Fongemie, Peter Juel Jensen & Patrick Jensen - for the best group work I have ever experienced.

Jesper Bo Sembach Christensen - for an extraordinary capability to learn and process data, as well as being a top notch manager.

Allan Riordan Boll - for being one of the most talented, innovative and kind Software Development Engineers I have ever met.

O Introduction

-LARS & ALEXANDER

Offering robust digital security is crucial in a modern society. Security concepts pervades the modern world in ways not readily apparent and as the world moves towards an ever increasingly digital world, the deployment, testing, understanding and auditing of IT security components become ever more crucial.

One of these components, is the *cryptographic hash algorithm* which is the focus of this report. In particular, the *Secure Hash Algorithm 1* commonly written as **SHA-1**¹ will be examined as it was deprecated December 31st 2015 by leading global tech companies such as Microsoft² and Google³, with the European research and education network TERENA/Géant following suit⁴.

The strength of **SHA-1** has been weakened through the years⁵, which is why it is important to ask the question: What are the consequences of not deprecating **SHA-1**?

As a Cryptographic Hash function, **SHA-1** is supposed to possess a set of mathematical properties which are:

- 1) Collision resistance: Infeasible to generate two identical hash values (from different inputs)
- 2) Pre-image resistance: Infeasible to derive the input from a hash
- 3) 2nd pre-image resistance: Infeasible to find a second input that has the same hash as another chosen input

A hash function with these properties can in turn be used to achieve these cryptographic building blocks:

- a) Data integrity – No change in a message without the hash changing.
- b) Authenticated data integrity – The last change done to the message was the author.
- c) Non-repudiation – An author cannot deny being the author.

There are three different collision types.

- Matching **RANDOM** data with **RANDOM** data. [general | random on random]

With the next two compromising 3), a), b) and c) from above:

- Matching **SPECIFIC** data with **RANDOM** data [2nd pre-image | specific on random]
- Matching **SPECIFIC** data with **SPECIFIC** data [2nd pre-image | specific on specific]

Throughout the report these have been named in accordance with the text used in the square brackets.

¹ 3rd and Jones, "RFC3174 - US Secure Hash Algorithm 1 (SHA1)."

² "SHA1 Deprecation Policy - Windows PKI Blog - Site Home - TechNet Blogs."

³ "Intent to Deprecate: SHA-1 Certificates - Google Groups."

⁴ "TERENA> News> TCS Certificate Service Responds to SHA Security Update."

⁵ Stevens, "Cryptanalysis of MD5 & SHA-1."



O.1 What is a hash?

-LARS & ALEXANDER

Following is a short explanation on hashing:

In order to detect modifications to electronic documents and insure integrity of data, *Digests*, **M**essage **A**uthentication **C**odes(**MAC**) or *hashes* are used to uniquely identify contents of a document, file or program. When transmitting data, along with its hash value, an extra layer of security is added against accidental or malicious modifications, since the message would not match the hash any more. Adding a public key signing step to this process turns it into a signature algorithm, allowing content to be authenticated as originating from the holder of the signing key, since only the holder could produce the **MAC**.

As long as proper key management is in effect and the encryption and hash algorithms are of suitable strength, creating another document, code or file with that same value should be infeasible. However, should the hash function not be strong enough, there are significant ramifications.

Like a car license plate there must not be two that are identical, otherwise a wrong person could be fined, and tied to criminal activity in the case of a falsified license plate.

A falsified hash on the other hand has way larger consequences, from impersonating a bank, train control center or the European Commission to telling a computer that malicious code indeed is an official Apple OSX Update. Another case for hashing is non-repudiation; proving that an action, decision or payment, was made by one specific legal entity, which is done by showing, that one and only one person had access to the specific key used, while also confirming timestamps⁶.

In a time with more and more electronic devices entering our homes and critical government infrastructure, the replacement of official firmware code with a malicious version having a backdoor, yet with the same identification code (hash) is a real and serious threat⁷.

What was thought to be a confidential digital conversation with an authenticated person, could turn out to be wiretapped or with a completely different entity, which is why it is fundamentally important to review the continued suitability of **SHA-1** as a cryptographic hash function.

For this reason, this thesis will focus on uncovering areas of application of cryptographic hash functions, with a focus on **SHA-1**. This will be the foundation for understanding the consequences of what could happen, should it be proven that **SHA-1** does not live up to the fundamental criteria.

By studying previous incidents, an estimation can be made of the potential consequences, and estimate some of the economic consequences as well as impact on industry and internet infrastructure.

Due to the high initial investment costs and long life-cycle, the railway sector will be investigated as well as key government infrastructure on national and European plan.

The goal of this thesis is to estimate how likely such an attack is with the hardware currently available today, using the 267th fastest computer made by man⁸ as a test platform.

Updating the 2012 estimates by Bruce Schneier⁹.

To do this estimate a custom **SHA-1** (brute)forcer application has been developed in order to evaluate the probability of a 2nd pre-image attack against a digital certificate, owned by the European commission (specific on specific collision). Using the **HPC** application, the certificate meta-data will be repeatedly modified and its hash value will be generated anew, in an attempt to find an identical **SHA-1** hash to the original, such that the

⁶ Itoh et al., "Forgery Attacks on Time-Stamp, Signed PDF and X.509 Certificate."

⁷ "Researchers Hijack Printer Using Malicious Firmware Update."

⁸ "TOP500 Supercomputer Sites | 267."

⁹ Schneier, "When Will We See Collisions for SHA-1? - Schneier on Security."



legitimate and forged certificate generate the same hash value when tested by a 3rd party, meaning the validity of the forged and the original controlled by the European Commission will be the same.

An alternative source of **SHA-1** values is explored, specifically the **BitTorrent** Network which predominantly builds upon **SHA-1** as an integrity mechanism, making it a possible candidate for pre-generated digests as well as a prime target for any attacks stemming from a weak hashing algorithm (specific on random collision).

The theoretical foundation, coupled with the experimental results of this report is used to provide an evaluation on the strength of the **SHA-1** function with the aim of trying to reevaluate Bruce Schneier’s estimate that **SHA-1** will not be broken before 2018¹⁰ and Stevens’s estimate of early autumn 2015¹¹ which is the overarching goal of this report.

O.2 Report Structure

-LARS

The first section of the report deals with theory, providing generic information to help understanding this report covering from Signing to Disclosure,

The second section is dedicated to applying the theory to the topics spanning from Apple Update Service to Shamir Secret Sharing.

As illustrated below:

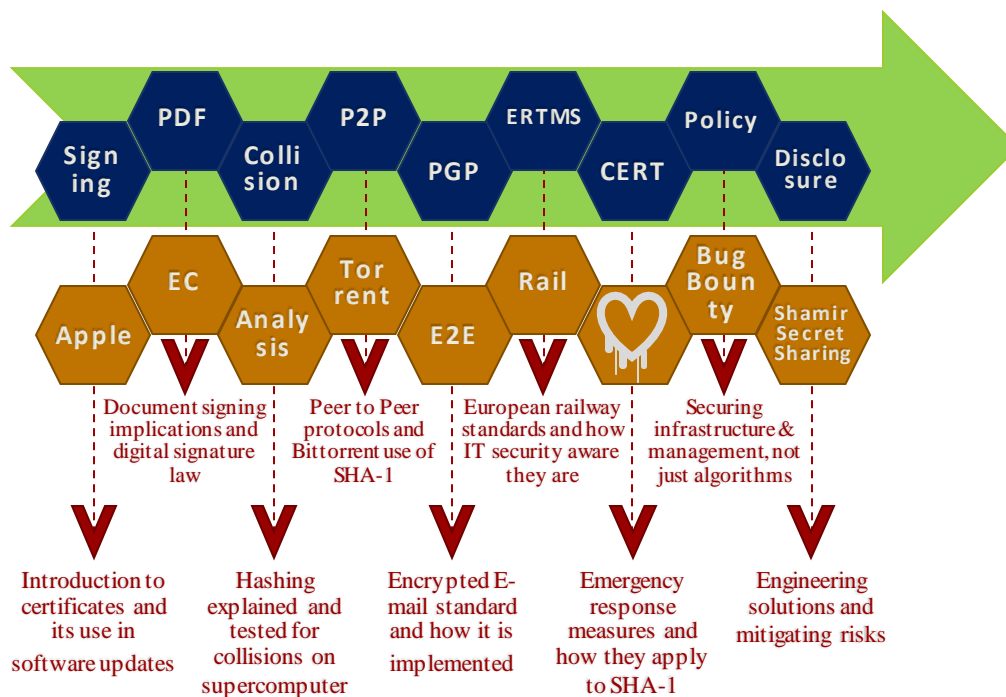


Figure 1 Graphical reading guide. Going from left to right as signified by the green arrow, the top blue row contains the overarching topics of the theoretical parts of the report and the bottom orange row contains the topics of the analysis. The Red arrows show themes that transcend the report, by Lars Embøll

For ease of reading **Abbreviations** and **Terms** that can be found in the end of this report (chapter 19, pages 111-114) are highlighted throughout the text.

¹⁰ Ibid.

¹¹ "The Shappinging."



PART 1 THEORY

-LARS

Leonard Nimoy famously said¹²:

“When you eliminate the impossible, whatever remains, however improbable, must be the truth.”, concluding that scientists should investigate errors to find causality.

This chapter will briefly touch and outline the theory that is used in later parts of this report.

It is meant as a short introduction and may be redundant for some readers, hence this is structured in a way that it should be possible to look up while reading the report sections where these topics will be referenced.

Confidentiality, integrity, authenticity and availability are core security aspects needed by any company in the information age. The security of many systems rely on the axiom that it is infeasible to find two different messages with the same hash, hence it is of the uttermost importance to investigate where they are used and the effects they have on those 4 core aspects and the system as a whole.

1 Common Cryptographic Hash Algorithms

-ALEXANDER

Cryptographic hash algorithms are a special class of hash algorithms, with specific properties such as general collision resistance (it is infeasible to find two different messages with the same hash value), pre-image resistance (It should be infeasible to generate a message such that its hash matches a previously chosen hash) and 2nd pre-image resistance (finding a second message with the same hash as a known message should be infeasible)¹³.

This section will not detail the construction of individual hash algorithms, but will instead focus on them from a black-box perspective, with the knowledge of existing attacks taken into account as well as the applications they are best suited for.

1.1 MD5

-ALEXANDER

MD5 is by now largely considered broken in cryptographic contexts.

It uses a digest space of 128 bits, and was introduced in 1992, where this was a respectable size. In 1996 attacks against it were severe enough that it was recommended to not use it for cryptographic means anymore.

From 2005 and forward, 2nd pre-image collision attacks could be performed in a couple of hours against **MD5**-based X.509 certificates with a standard laptop¹⁴.

Beyond cryptanalysis based attacks, the key space of 128 bits is today considered too small for the algorithm to be secure against even a pure birthday attack^{15 16} by supercomputers.

¹² Doyle and Kerr, *The Sign of Four*.

¹³ Pfleeger and Pfleeger, *Security in Computing*.

¹⁴ Klima, “Finding MD5 Collisions-a Toy For a Notebook.”

¹⁵ Stevens et al., “Short Chosen-Prefix Collisions for MD5 and the Creation of a Rogue CA Certificate.”

¹⁶ “Microsoft Word - MD5 Collisions Whitepaper.doc - wp.MD5_Collisions.en_us.pdf.”



1.2 SHA-1

-ALEXANDER

SHA-1¹⁷ was introduced in 1995 as a replacement for **SHA-0**, which in turn was a replacement for **MD5**, it features a digest space of 160 bits. While no attacks against the full **SHA-1** function has been performed yet, it is estimated to be within the current or the next generation of hardware capabilities, and for that reason it is considered deprecated and all cryptographic use of it should be phased out^{18 19}.

When applying the pigeonhole principle (see chapter 3 Collision Probability, page 27), the amount of guesses needed to approach a 50% chance of general collision is 2^{80} , and while this is a significant amount of guess, recent advancements have brought the chance of a general collision down to 2^{61} . Furthermore, if a specific initialization vector is chosen (it represents the **SHA-1** internal state between input blocks), the strength of the function is brought down to 2^{50} .

This serves to illustrate that under the right conditions, the strength of **SHA-1** can be significantly less than advertised.

1.3 SHA-2

-ALEXANDER

The successor to **SHA-1** is **SHA-2** and it is, opposed to **SHA-1**, a family of hash algorithms with a variable digest space depending on the version in use. What is common for all versions is that they have more than 220 bits in the digest space, with the longest version featuring up to 512 bits.

Attacks have been found however, which significantly lowers the amount of secure bits²⁰ for the entire family of **SHA-2**, but it is still harder to produce any type of collision for **SHA-2** than **SHA-1**.

1.4 SHA-3

-ALEXANDER

The newly released (august 5 2015) algorithm **SHA-3**, and while it shares the SHA name, it is functionally not related to **SHA-1/2**²¹.

It was released as the result of a five-year competition for the next generation of SHA, and the winning algorithm was chosen for better performance than the **SHA-2** family and due to it having another, but proven architecture, which did not suffer from attacks already known in the **SHA 1/2** family.

Like **SHA-2**, **SHA-3** implements a family of algorithms, which are based around the central algorithm with a modulo of its output, constructed to match that of **SHA-2**. This as a consequence means it features the same amount secure bits as **SHA-2**, however it features none of the known attacks. Also the internal algorithm can be tuned to provide much larger digest lengths, expanding its potential lifetime.

¹⁷ Dang, "Secure Hash Standard (SHA-1) NIST FIPS 180-4," 1.

¹⁸ Andrews, "The Cost of Creating Collisions Using SHA-1," 1.

¹⁹ Karpman, Peyrin, and Stevens, "Practical Free-Start Collision Attacks on 76-Step SHA-1," 1.

²⁰ "286.pdf."

²¹ US Department of Commerce, "NIST Selects Winner of Secure Hash Algorithm (SHA-3) Competition."



2 Risk Management Theory

-LARS

Systematically finding errors in a complex environment is impossible, leading to the creation of tools trying to section complex systems into fewer complicated sections, trying to parameterize hazards, consequences and barriers in order to better handle them.

Risk management methods can be used in:

- The design phase of a project to mitigate risk and form acceptance levels.
- In existing protocols to identify faults and outcomes.
- Comparing and rank ordering risks.

As the terminology is new to a large percentage of people dealing with software the image below illustrates commonly used terminology detailing the difference between *Hazard Identification*(green), *Risk Analysis*(red) and *Risk Assessment*(yellow) in a flow diagram detailing the process.

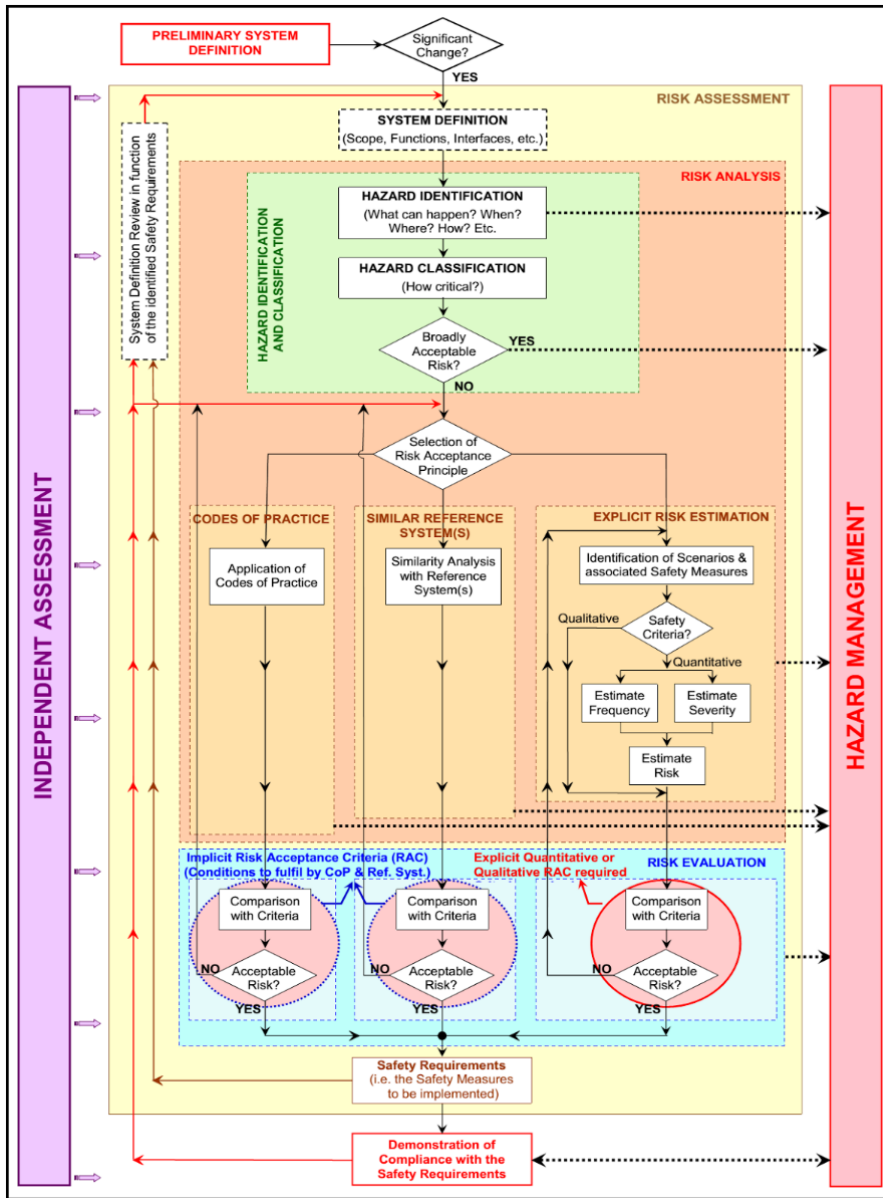


Figure 2 Commonly used terminology detailing the difference between Hazard Identification, Risk Analysis and Risk Assessment by ²²

²² Jovicic, "ERA Guide for Application of the Common Safety Methods on Risk Assessment."



As the goal of this project is to quantify the strength of hash functions *Risk Analysis tools* will be used for *Hazard identification* and comparison with existing threats.

Prominent tools used in *Hazard identification* are **FMEA** (Failure **M**ode and **E**ffect **A**nalysis)²³ searching for triggers, following them down to a consequence; a bottom-up inductive method is **HAZOP** (**H**AZard and **O**Perability analysis)²⁴ also seen in **Event trees**.

These tools formalise the process of finding vulnerabilities in critical systems by exploring possible outcomes and are recommended in the process of finding software vulnerabilities.

2.1 Bathtub & Rocking Boat

-LARS

New systems and practices are known to cause errors due to an unfamiliar environment, unexpected loopholes and changed management practices. This leads to the conservatism of using older familiar systems, but as the following figure describes it will in turn lead to complacency and an unfounded assumption that nothing can go wrong, because no error has happened in a long time, leading to rules being bent and not enforced to their original intent.

The bathtub curve is normally used to illustrate wear and tear of physical components but can easily assist the *rocking boat* mentality of slacking on safety rules when there have been no errors for a generation of employees. For companies with a high employee turnover / churn, a generation of employees with no memory of errors can be as low as few years.

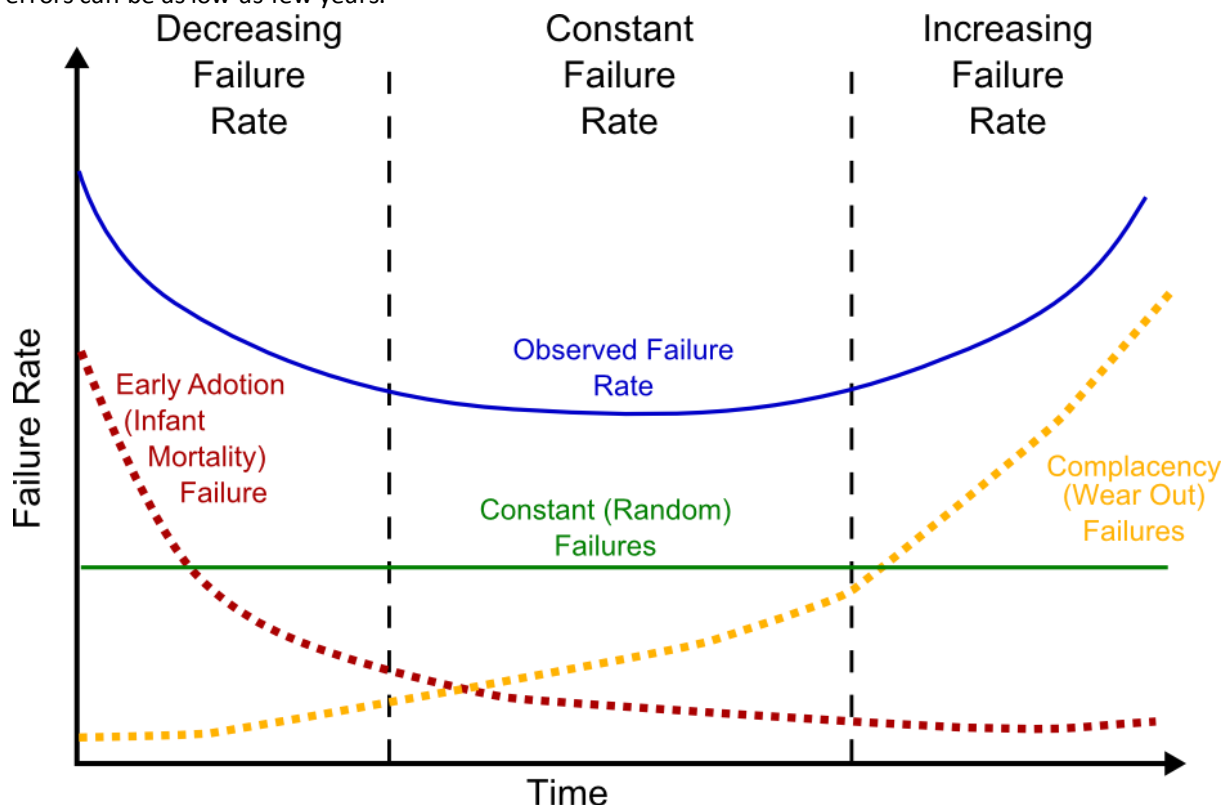


Figure 3 Bathtub principle drawing, by Lars Embøll, derivative of public domain work

²³ Apollo Reliability and Quality Assurance Office, "Procedure for Failure Mode, Effects and Criticality Analysis (FMECA)."

²⁴ Imperial Chemical Industries, Chemical Industries Association, and Chemical Industry Safety & Health Council, *A Guide to Hazard and Operability Studies*.



The bathtub hazard function is a simplification used to easily explain the constituent elements that a typical hazard function (blue line) consists of:

- infant mortality / early adoption problems (red)
- consistent random errors (green)
- wear out / complacency (yellow)

For a more evidence based and mathematical correct hazard function ²⁵ provides a better methodology for producing accurate graphs based on observed data.

2.1.1 Rocking Boat Principle

-LARS

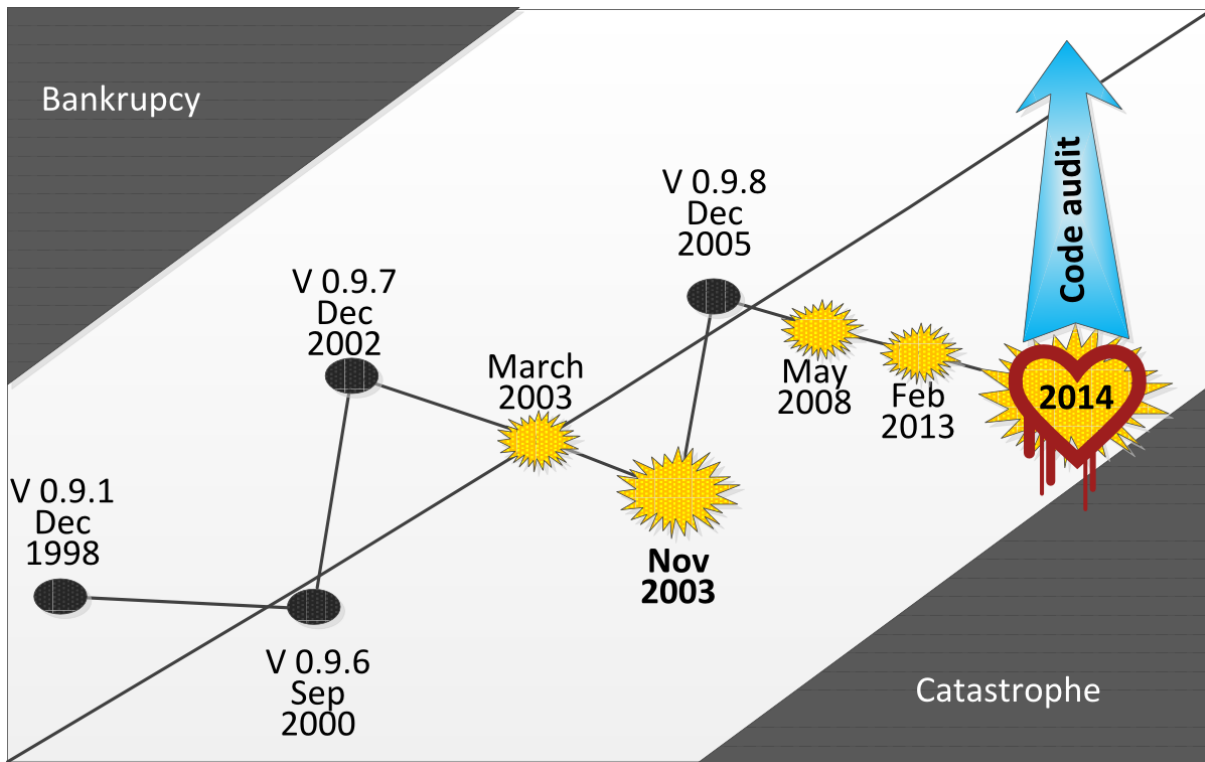


Figure 4 Rocking boat principle with OpenSSL example, by Lars Embøll, derivative of ²⁶

The rocking boat principle is a complacency effect as seen above the security & funding is increased when an incident has happened rather than evenly over time.

2.2 Risk Statistics

-LARS

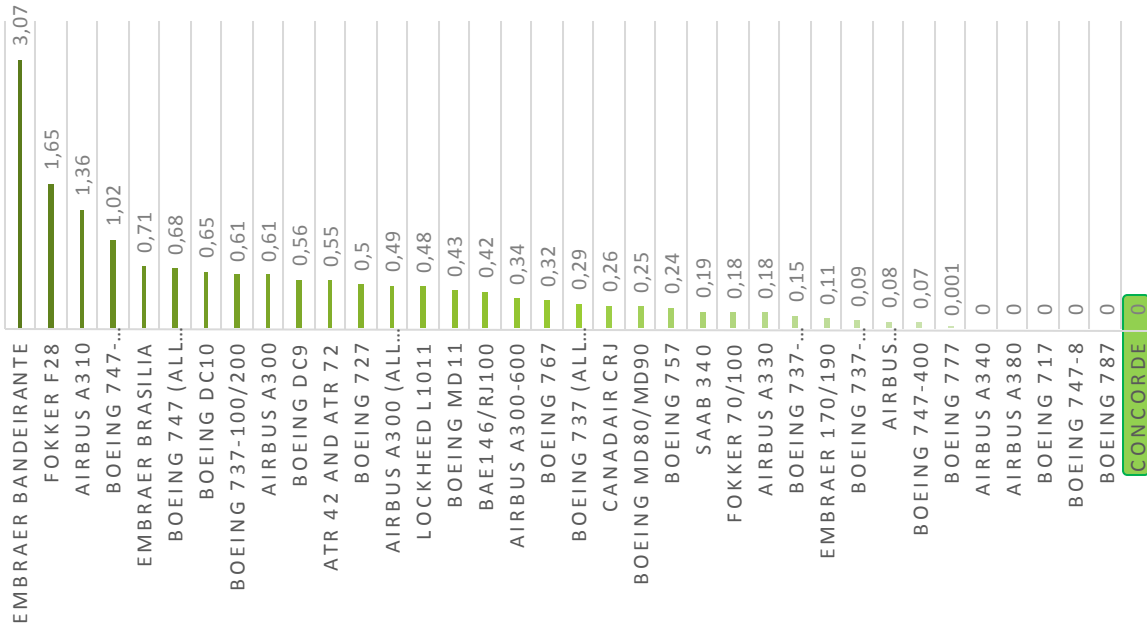
Risk is probability multiplied by consequences, meaning that low probability, high consequence events have a high impact on averages, as in the case of the Concorde:

²⁵ Klutke, Kiessler, and Wortman, "A Critical Look at the Bathtub Curve."

²⁶ Reason, *Managing the Risks of Organizational Accidents*.



FATAL CRASH RATES PER MILLION FLIGHTS 24 JULY 2000



FATAL CRASH RATES PER MILLION FLIGHTS 25 JULY 2000

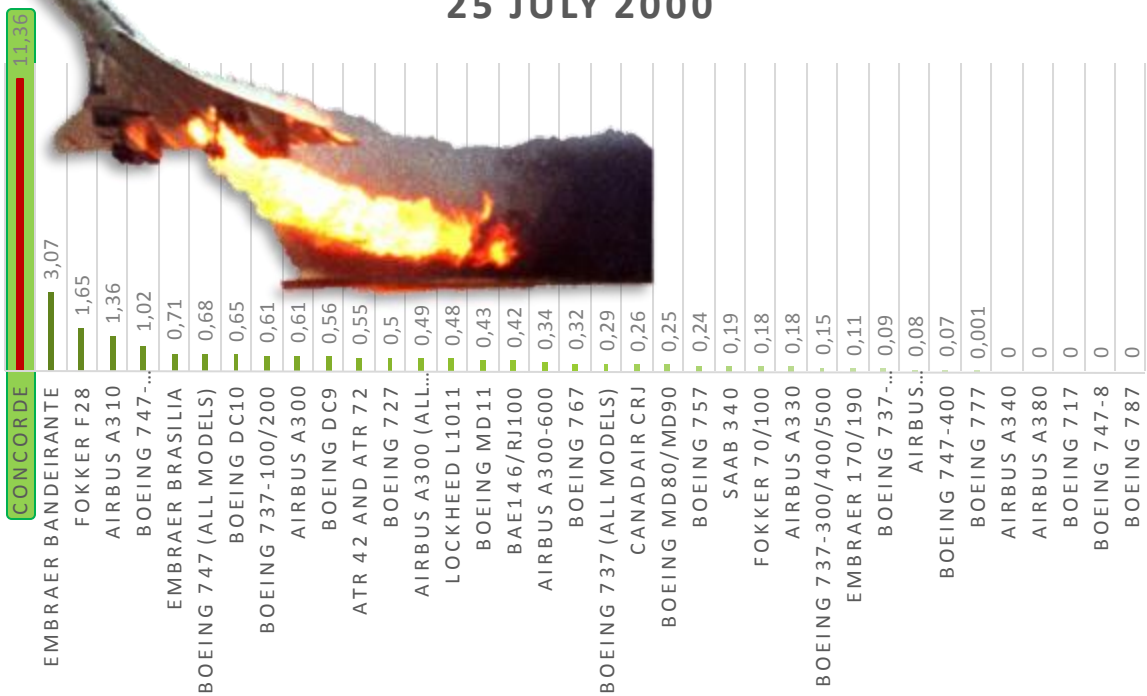


Figure 5 Aviation safety, by Lars Embøll, data from ^{27,28}.

Picture: ©Tashihiko Sato, Associated Press Air France Concorde flight 4590, 109 deaths

²⁷ "Aviation Safety Network > ASN Aviation Safety Database > Aircraft Type Index."

²⁸ "Fatal Plane Crash Rates by Model."



Where the fatal crashes per million flights goes from zero to almost 4 times worse than the airplane with the 2nd highest amount of crashes.

2.2.1 Near Misses & Bug Reports

-LARS

There is a plethora of ways to deal with incident reports, leading to academic papers trying to classify, weigh and compare the methods. This chapter describes the widely used incident pyramid and the theory of incident report handling.

2.2.1.1 The Incident Pyramid

-LARS

The hypothesis behind the incident pyramid is that the number of fatal accidents, reported incidents, near misses and safety rule violations are correlated. Since 1931²⁹ an estimate for this has been sought, with an estimate from 2011 being shown below:



Figure 6 Statistics of incident to fatality ratio, by Lars Embøll, data from ³⁰

This can be extended with estimates from a 2003 ConocoPhillips study³¹, that gives the following numbers:

²⁹ Heinrich, *Industrial Accident Prevention*.

³⁰ Collins, "Heinrich's Fourth Dimension."

³¹ Freibott, "Sustainable Safety Management."



Figure 7 Incident Pyramid, including at-risk-behaviours, by Lars Embøll, data from ^{32,33}

2.2.1.2 Responsible Disclosure in a Risk Assessment Perspective

-LARS

While the incident pyramid creates an estimate for the average distribution of accidents, the proportion of incident reports relies on the management culture of the workplace.

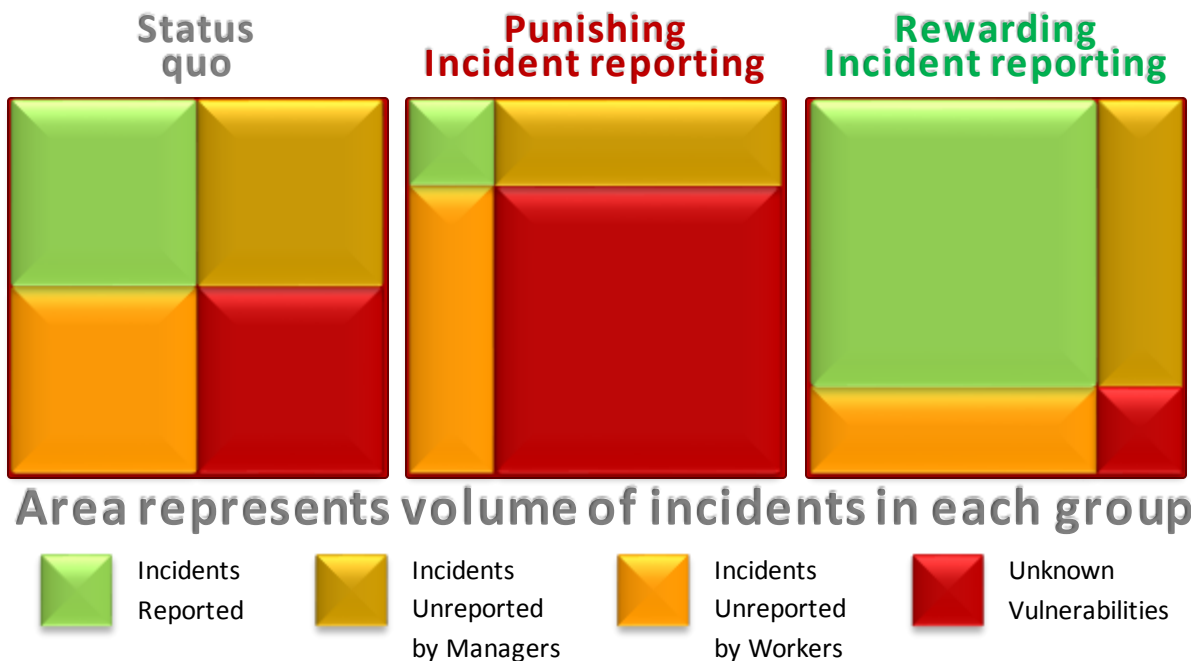


Figure 8 Share of vulnerabilities known and reported based on whistleblower policy, by Lars Embøll derivative of ³⁴ The risk management theory shown above relates to workplace accidents, but also applies to cyber and on-site security.

³² Collins, "Heinrich's Fourth Dimension."

³³ Freibott, "Sustainable Safety Management."

³⁴ Borg, "Predictive Safety from Near Miss Hazard-Reporting."



A hypothetical example to illustrate this principle could be a case where there is a new employee, sadly the secretary is on holiday, so the new employee cannot get a key.

A long term employee mentions that locks are poorly shielded and that they can be opened by jamming a business card in between the frame and the door.

20 days passes and management introduce a program to increase and encourage submitting incident reports of near misses not only security breaches causing a loss.

The employee then has two options:

1. Not report it and risk an intruder using the same vulnerability
2. Report the doors being easy to open and risk getting fired for having misused this for 20 days

Given that the employee reports the issue, the manager also has two options:

1. Punish the employee for not having reported it earlier
2. Reward the employee for the report and fix the issue

It seems counter-intuitive to reward employees for their bad behaviour, but following the easy 1st choices lead to more open vulnerabilities.

The easy management choice is to punish breaches of company rules, thus making the precedence that people who file reports of issues that have been known for a long time, will be actively dis-incentivised to report incidents (centre illustration).

While company rules, and the law in principle, should be followed the company will have less knowledge of vulnerabilities and be open for attacks, or in the case of near miss work incident reporting have a larger risk of fatal accidents.

2.3 Risk Acceptance

-LARS

While previous chapters have focused on explaining risk and risk reduction through general mitigation techniques, this chapter will explore international standards and their methods to parametrise risk for comparison.

Risk acceptance, unlike direct financial impacts, is not finite and countable.

A way to judge risk acceptance is how much agency the subject has and the degree of culpa from the acting part. While the consequence is the same from a fatal rock climber accident and a murder, the lack of agency leads to a higher perceived cost for society and a *willingness to pay* that is larger for investigating and avoiding murders than rock climbing accidents.

Compromised IT security often have an impact on a lot of people due to the monoculture of programs/OS fostered by positive externalities and economy of scale.

Hence why the price for executing known attacks are extremely low compared to the costs it incurs on the target(s).

Given that economy of scale is a strong economic force, standardisation pays off, once a service or platform has reached critical mass the marginal cost for new users decrease for the system owners, while strengthening the positive externality for other users joining around the same platform. But with a lot of users on a platform(monoculture) a vulnerability to that platform gives access to a lot of users(attack surface).

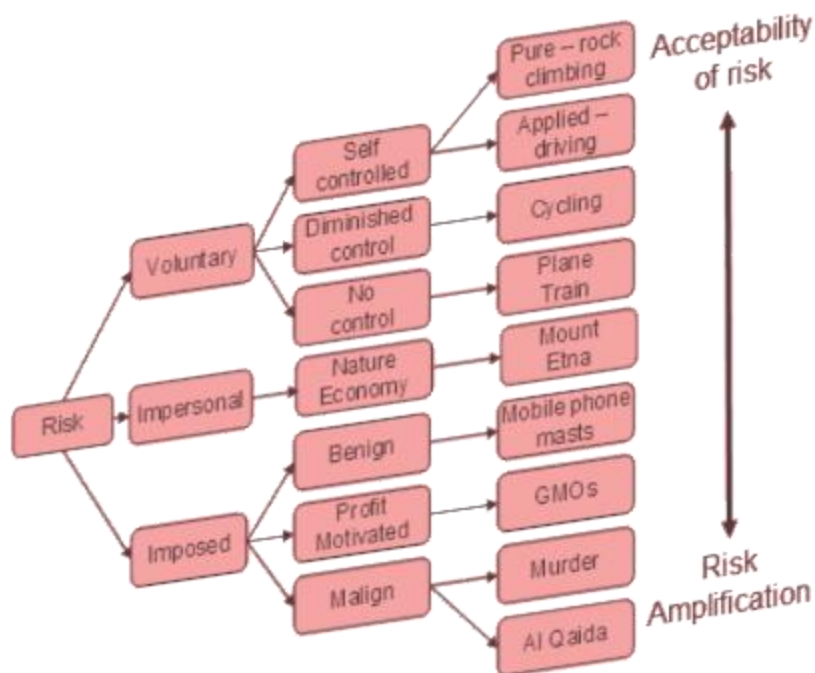


Figure 9 Risk acceptability by ³⁵

Drivers assume a great deal of responsibility by being the agent in control of the vehicle both in respect to handling and maintenance, compared to boarding public transportation where the traveller has no direct influence on safety.

But culpa is not the only factor, medical response and hospitals are built to cope with an Erlang distribution of injuries³⁶, accommodating for one car induced injury per million capita each day nationwide³⁷, rather than hundreds of injuries from a train or airplane accident in a local area.

Lastly there is a big difference between an identified individual and a statistical life. Thomas C. Schelling puts it well in the following quote:

“Let a six-year-old girl with brown hair need thousands of dollars for an operation that will prolong her life until Christmas, and the post office will be swamped with nickels and dimes to save her. But let it be reported that without a sales tax the hospital facilities of Massachusetts will deteriorate and cause a barely perceptible increase in preventable deaths-not many will drop a tear or reach for their checkbooks.”

- ³⁸ PAGE 115

These are reasoning for ambiguity aversion³⁹ and the difference in valuation of a casualty depending on the degree of culpa, number of people injured at the same time and identification to a population subgroup. This is without accounting for the epistemic uncertainty in the Danish evaluation method⁴⁰ described in chapter 2.4.2 Epistemic uncertainty in Danish VSL, page 27.

³⁵ Adams, “The Economics and Morality of Safety Revisited.”

³⁶ A. M. de Bruin, “Dimensioning Hospital Wards Using the Erlang Loss Model. Ann Oper Res.”

³⁷ Statistics Denmark, “Traffic Accidents with Injuries.”

³⁸ Schelling, *Choice and Consequence*.

³⁹ Treich, “The Value of a Statistical Life under Ambiguity Aversion.”

⁴⁰ Danish Ministry of Transport and COWI, “Rapport om værdisætning af transportens eksterne omkostninger.”



2.3.1 Railway Safety

-LARS

In regards to security and safety the railway historically has had a conservative and high safety approach leading trains to be one of the safest modes of transportation.

With infrastructure and rolling stock often lasting decades, it is interesting from a security perspective, as this long operational time will have to be taken into consideration going from electro-mechanic systems that can have proven safe states to a field of IT security resting on *computational hard* problems, where some problems during the course of 5-10 years have been downgraded to feasible⁴¹.

This chapter is predominantly based on publicly available information, using standards and reports such as the censored *ERTMS IT Security Threat identification, Risk Analysis and Recommendations*⁴², due to the difficulty of obtaining information within the railway sector. The domain seems interested in risk analysis results, but reluctant to provide input beyond pointing to the list of **ERTMS** standards.

Based on the open source repository of the ERTMS Formal Specs⁴³, the only trace of **SHA-1** was that since April 10th 2015 **MD5** was replaced with **SHA-1** in the installation software⁴⁴(**LINE 177**).

In 2011 a safety analysis noted the use of **DES** within the *GSM-R* standard, suggesting a replacement with **AES**⁴⁵. The implementation of triple **DES** is described in ⁴⁶ **ANNEX E**, with a summary in chapter **7.2**.

2.3.2 SIL & IEC 61508

-LARS

While most standards and protocols dealing with IT are trivial, *IEC 61508* has a wide and complex range of specifications and requirements for documentation more akin to “what is the value of a human life?” than “number of bits in the key”

A specific example from ⁴⁷ **PART 3** being: “6.2.3 *Software configuration management shall c) maintain accurately and with unique identification all configuration items which are necessary to meet the safety integrity requirements of the E/E/PE safety-related system.*”

Displaying how vague wording is used rather than specific examples for implementation, making it complex to implement compared to **NIST** standards specifying what algorithms and key lengths to use⁴⁸.

A main component of *IEC 61508* is the notion of security and safety not being better than the most vulnerable component, as illustrated in the previous subchapters, as well as the **SIL** 0-4 mentioned in chapter 2.3.1 Railway Safety, page 21.

What was not mentioned though was the perspective of dealing with failure rates less than 1 in 10'000 or once each 100'000'000 hours for **S**afety **I**ntegrity **L**evel 4, 10^8 is 11'416 years.

As the system has to be proven to be within the specified **SIL** level there needs to be a buffer accounting for uncertainties, but also cutting costs by not being right below the upper bound of a **SIL** level, as that increase production cost, hence the mean is a good estimate for actual components.

⁴¹ Stevens et al., “Short Chosen-Prefix Collisions for MD5 and the Creation of a Rogue CA Certificate.”

⁴² KPMG IT Advisory, “ERTMS IT Security Threat Identification, Risk Analysis and Recommendations PUBLIC VERSION.”

⁴³ “ERTMS Solutions | ERTMSFormalSpecs - Open Source - ERTMS Solutions.”

⁴⁴ “ERTMSFormalSpecs InnoInstaller5/whatsnew.htm.”

⁴⁵ Mária Franeková, “Safety Analysis of Cryptography Mechanisms Used in GSM for Railway.”

⁴⁶ “EuroRadio FIS - SUBSET-037.”

⁴⁷ International Electrotechnical Commission, “IEC 61508 Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems.”

⁴⁸ Barker et al., “Recommendation for Key Management SP 800-57 Part 1: General Revision 3.”



Safety Integrity Level (SIL)	Average probability of a dangerous failure on demand of the safety function (PFDavg)	Average probability of a dangerous failure of the safety function [h^{-1}] (PFH)	Mean time between failures of a dangerous failure of the safety function [years] (MBF)
4	$\geq 10^{-5}$ to $< 10^{-4}$	$\geq 10^{-9}$ to $< 10^{-8}$	57'078
3	$\geq 10^{-4}$ to $< 10^{-3}$	$\geq 10^{-8}$ to $< 10^{-7}$	5'708
2	$\geq 10^{-3}$ to $< 10^{-2}$	$\geq 10^{-7}$ to $< 10^{-6}$	571
1	$\geq 10^{-2}$ to $< 10^{-1}$	$\geq 10^{-6}$ to $< 10^{-5}$	57

To relate **SIL** levels with human history:

60'000 years ago is when humanity was confined to Africa⁴⁹ ~ **SIL 4**

5'000 years ago marked the foundation of Troy ~ **SIL 3**

564 years ago Christopher Columbus was born ~ **SIL 2**

70 years ago we had WWII ~ **SIL 1**

Planning for a system not to fail within the scope of humanity, not only the 5'000 years since the unification of ancient Egypt under the first pharaoh, but 10 times further back when man had a population of only 2'000 individuals⁵⁰, seems illogical and impossible, but puts things in perspective.

With a production run of one million, 60'000 years of run time can be experienced each 20 days of continuous use of the whole production run.

But it leads to uncertainty for low production runs, while there may be millions of cars, TVs and smartphones, trains are quite limited in their numbers.

SIL levels apply to systems, a car could be a system, sadly, it is also unclear what the scope of the **SIL** systems are; if a population of 60'000 cars having 1 failure each year on a brake is needed for **SIL 4** or if you just need 15'000 cars having a failure on one of their 4 wheel brakes to qualify for **SIL 4**.

2.3.3 European Railway Security

-LARS

The European railway is broadly sectioned in two groups: **TSI** and non-**TSI**.

TSI being **T**echnical **S**pecifications for **I**nteroperability.

Stretches of railway governed by **TSI** (part of the Trans European Network for Transportation or TEN-T) falls under **ERA** jurisdiction in order to ensure free flow of goods within the European Union(**EU**).

Part of this regulation set is the proposed harmonization of signalling standards:

ERTMS⁵¹ (**E**uropean **R**ail **T**raffic **M**anagement **S**ystem)

The responsibility for the IT-security of **ERTMS** fall upon **ENISA** (**E**uropean **N**etwork and **I**nformation **S**ecurity **A**gency) though⁵².

⁴⁹ A Family Tree for Humanity.

⁵⁰ Ibid.

⁵¹ "Set of Specifications #2 (ETCS Baseline 3 and GSM-R Baseline 0)."

⁵² European Railway Agency Corporate Management and Evaluation, "FW: Information Request Form - Nielsen (Dec 2)."



2.4

2.4 Risk Classification

-LARS

Standards need to be able to quantify risk, splitting it up in its components of probability and consequence. While **IEC61508** details probability to a great extent, but only has a weak bond to specific consequences for **SIL** levels, that can only be found in the annex C of **IEC61508**⁵⁴ (**PART 5**) referencing the **ALARP** principle.

ALARP relates to the cost of a lost human life.

So with a valuation of a human life, a monetary value can be directly linked to a **SIL** and hence give an indicator of the damage a cyber-attack should incur in order to require precautionary measures to the extent of **SIL 4**, with the interesting question if readily available **SHA-1** general or 2nd pre-image collisions is of that magnitude.

2.4.1 The money value of a man

-LARS

Each year, European countries are required to report their national estimate of “**V**alue of **P**reventing a **C**asualty”, **VPC** to the **E**uropean **R**ail **A**gency (**ERA**) due to the “*Commission Directive 2009/149/EC of 27 November 2009 amending Directive 2004/49/EC of the European Parliament and of the Council as regards Common Safety Indicators and common methods to calculate accident costs*”⁵⁵, specifically **R11**⁵⁶ and **R16**⁵⁷ with the Danish Value of Preventing a Fatality being 2’839’534.88372€⁵⁸ in 2014, though it has a high degree of uncertainty⁵⁹, it is the official value for Denmark⁶⁰ (§ 79 stk 2).

In order to understand the number and how it translates into monetary value it is important to know the models used to derive the value, as they are very different and hence not directly comparable.

With some economists using the Human Capital(**HK**) approach devised by Dublin & Lotka⁶¹ from the 1930s for quantification of risk.

Below is a brief summary on methodologies for the Money value of a man:

2.4.1.1 Human Capital (HK)

-LARS

In 1954 Reynolds writes “The Cost of Road Accidents”⁶² which mentions:

“The occurrence of road accidents inflicts a burden on the community which may be considered in two parts.

(i)The pain, fear, and suffering imposed by the occurrence, or the risk of occurrence, of road accidents. These are considered of great importance in a society that values human life and human welfare.

(ii)The more concrete and ascertainable burdens in the form of the net loss of output of goods and services due to death and injury and the expenditure of resources necessary to make good the effects of accidents, e.g. medical expenses, vehicle repairs and costs of administration.

⁵⁴ International Electrotechnical Commission, “IEC 61508 Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems.”

⁵⁵ *Commission Directive 2009/149/EC.*

⁵⁶ “Common Safety Indicators Reported by the National Safety Authorities - R11 - National Value of Preventing a Fatality - Denmark 2006-2014.”

⁵⁷ “Common Safety Indicators Reported by the National Safety Authorities - R16 - Fall Back Value of Preventing a Fatality - Denmark 2006-2014.”

⁵⁸ “Common Safety Indicators, Denmark 2014, Version 1, Validated (R11).”

⁵⁹ Danish Ministry of Transport and COWI, “Rapport om værdisætning af transportens eksterne omkostninger.”

⁶⁰ “Jernbanelov - Retsinformation.dk.”

⁶¹ Dublin and Lotka, *The Money Value of a Man.*

⁶² Reynolds, “The Cost of Road Accidents.”



For a variety of reasons it is beyond the competence of the economist to assign objective values to the losses suffered under (i) and this paper is therefore confined to the estimation of the burdens listed under (ii). "

While the evaluation of factors in (i) are clearly mentioned as missing, the values and methods of obtaining (ii), using the term "Human Capital" (**HK**) have been used for decades as the only value of asserting the cost for society regarding risk of casualties up into the 1960s⁶³, with the methodology being used until 1977 by the Danish **N**ational **S**afety **A**uthority (Trafikstyrelsen)⁶⁴.

Following the **HK** approach, there is no incentive to help people who are unable to contribute financially to society such as elderly and handicapped citizens, actually there is an incentive to lessen the safety levels of those groups, using the money on the labour force instead.

This decreased prioritisation of safety for the population not contributing positively to the **GDP** can easily lead to the "dead-anyway" effect⁶⁵.

It is this absence of (ii) that leads to the next development; the **V**alue of a **S**tatistical **L**ife (**VSL**).

2.4.1.2 Value of a Statistical Life (VSL)

-LARS

In their T430 report (PAGE 30)⁶⁶ The British **R**ail **S**afety and **S**tandards **B**oard (**RSSB**) defines **VSL** as:

"A willingness to pay-based VPC is essentially the aggregate, across affected members of society, of individual willingness to pay for (typically very small) risk reductions which will on average prevent one fatality. What the VPC is most emphatically not is the "price of a life" in the sense of a sum that would compensate the typical individual for the certainty of his/her own premature death – for most of us no sum, however large, would serve this purpose." [edited to account for other abbreviation use of the RSSB]

VSL is in other words, the added value put on top of society's loss of **GDP** (**HK**), to account for human life being more precious than the net product contributed to society⁶⁷. This is akin to the appreciation and hence monetary evaluation of "preservation of green areas in cities" and "endangered wildlife" that have become part of the **C**ost-**B**enefit **A**nalysis (**CBA**) with the advance of **M**ultiple-**C**riteria **D**ecision **A**nalysis (**MCDA**) that are used in the railway and road sector.

⁶³ Hultkrantz and Svensson, "The Value of a Statistical Life in Sweden."

⁶⁴ COWI and Vejdirektoratet, "Trafikøkonomiske Enhedspriser for uheld - Alternative metoder til opgørelse af Velfærdstabet (Arbejdsnotat)."

⁶⁵ Pratt and Zeckhauser, "Willingness to Pay and the Distribution of Risk and Wealth."

⁶⁶ Rail Safety & Standards Board, "T430 Assessment of the Value for Preventing a Fatality Phase 1."

⁶⁷ Shogren et al., "Resolving Differences in Willingness to Pay and Willingness to Accept."

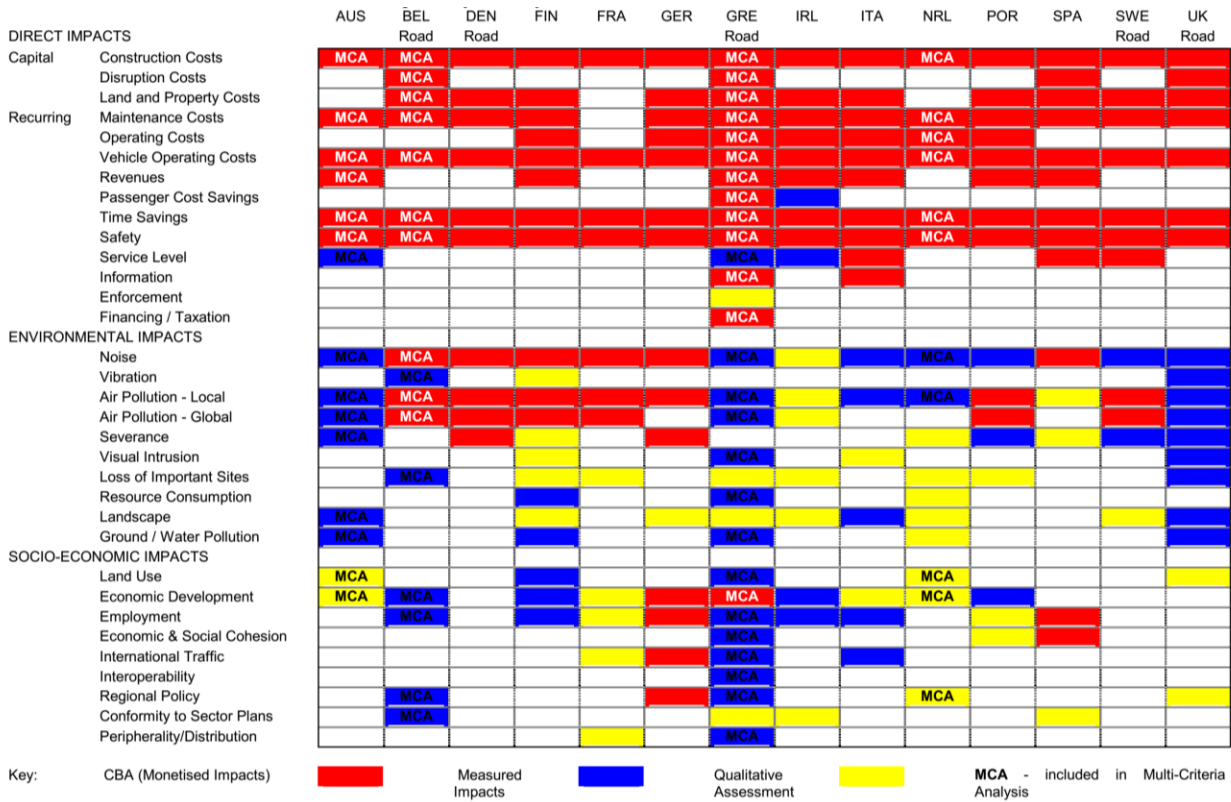


Figure 10 MCDA criteria used in various European countries ⁶⁸(PAGE 24)

The figure above shows how European countries vary in their use of components in a national infrastructure Cost Benefit Analysis. Values that cannot be measured on the free market as they are public goods, that have to be estimated via proxies in **R**evealed **P**reference or **S**tated **P**reference studies; so called “soft methods” marked in yellow above. Combined with economic terms such as time savings, construction and maintenance costs; so called “hard methods” marked in red and blue above.

Combining the soft values of **VSL** with the hard numbers of **HK** a more comprehensive method emerges: The **VPC**.

2.4.1.3 Value of Preventing a Casualty

-LARS

This method springs from the combination of asserting a value to human life and emotional suffering of the family, as well as accounting for the loss of **GDP** for society.

It is the method **ERA** / European Commission Directive 2009/149/EC⁶⁹ requires the member states to use.

Hence it is a Common Safety Indicator that has widespread use throughout the railway sector in Europe.

⁶⁸ EUNET / European Commission, “Socio-Economic and Spatial Impacts of Transport.”

⁶⁹ Commission Directive 2009/149/EC.

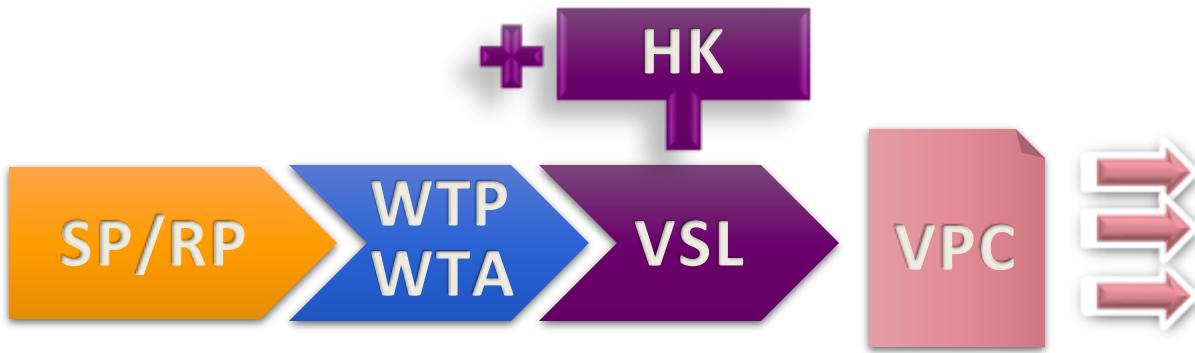


Figure 11 Basics of VPC: Stated/Revealed Preference surveys gives a Willingness To Pay/Accept leading to a Value of a Statistical Life, that combined with Human Kapital gives a Value of Preventing a Casualty

2.4.1.4 ALARP

-LARS

The **ALARP** principle stems from a valuation of life: The risk of causing death to an employee.

Starting with British mines, workers wanted their employers to improve the safety rather than just considering profit as a meter and to do so the British justice system and government needed a method to weigh costs of safety measures against that of the risk of a lost human life.

The result being the “**As Low As Reasonably Practicable**” methodology of the 1950s⁷⁰(PAGE 5).

It is this British **ALARP** principle that **IEC 61508** references, with the classifications seen below:

Table 1 Risk classification from **IEC61508**⁷¹(CHAPTER 5), based on **ALARP**⁷².

FREQUENCY	CONSEQUENCE			
	CATASTROPHIC	CRITICAL	MARGINAL	NEGLIGIBLE
FREQUENT	I	I	I	II
PROBABLE	I	I	II	III
OCCASIONAL	I	II	III	III
REMOTE	II	III	III	IV
IMPROBABLE	III	III	IV	IV
INCREDIBLE	IV	IV	IV	IV

NOTE 1 The actual population with risk classes I, II, III and IV will be sector dependent and will also depend upon what the actual frequencies are for frequent, probable, etc. Therefore, this table should be seen as an example of how such a table could be populated, rather than as a specification for future use.

NOTE 2 Determination of the safety integrity level from the frequencies in this table is outlined in Annex D in IEC 61508.

Interpretation of risk classes

RISK CLASS	INTERPRETATION
CLASS I	Intolerable risk
CLASS II	Undesirable risk, and tolerable only if risk reduction is impracticable or if the costs are grossly disproportionate to the improvement gained
CLASS III	Tolerable risk if the cost of risk reduction would exceed the improvement gained
CLASS IV	Negligible risk

⁷⁰ Rail Safety & Standards Board, “T430 Assessment of the Value for Preventing a Fatality Phase 1.”

⁷¹ International Electrotechnical Commission, “IEC 61508 Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems.”

⁷² Great Britain. Health and Safety Executive, *Reducing Risks, Protecting People*.



With the standard noting that:

“Frequent could denote an event that is likely to be continually experienced, which could be specified as a frequency greater than 10 per year. A critical consequence could be a single death and/or multiple severe injuries”

While vague, it gives enough information to classify cyber threats and with the conversion factor of **VPC** monetary damages can be extrapolated to desired **SIL** levels as well.

2.4.2 Epistemic uncertainty in Danish VSL

-LARS

The uncertainty stems from a report done by Kidholm in 1995, where 55% of the respondents took family into consideration⁷³ (PAGE 129, 149), as well as the yearly adjustment model that was updated in 2010, adjusting the figure to 15'000'000,00 DKK in 2007 prices⁷⁴.

Those are the reasons for a suggested sensitivity analysis of 300% and 33% by the ministry of finance⁷⁵ and why that while the official Danish valuation is 2'839'534,88372€ (ERA CSI R16) the amount of digits are misleading, they stem from a number being set to 15 million DKK with an uncertainty of a factor 3.

3 Collision Probability

-ALEXANDER

The primary strength of cryptographic hash functions is in its non-reversible nature, as well as the infeasibility of finding two inputs that produce the same output. It is however not impossible and this section is dedicated to exploring the probabilistic constraints governing hash functions and their collisions. The idealized hash methods under consideration in this section are treated as black-box function and any attacks against specific hash algorithms are ignored. All mentions of output size is in terms of total amount of values it can attain and not the amount of bits.

3.1 Hash Collisions

-ALEXANDER

When considering the idealized hash function $H_x(m_1) = d_1$, with an arbitrary input m_1 and output d_1 of that falls within the length x (the output can assume x distinct values), the probability of two non-equal input producing the same output is:

$$P(H_x(m_1) = H_x(m_2)) = \frac{1}{x}$$

This implies that as long as x is “sufficiently large” it is impractical to guess values of m such that they produce an equal result.

This is the fundamental theorem hash functions rely on and what is used to determine their strength.

⁷³ Kidholm, Odense Universitet, and Center for Helsetjenesteforskning og Socialpolitik, “Estimation af betalingsvilje for forebyggelse af personskader ved trafikulykker.”

⁷⁴ Willumsen, Jensen, and Hansen, “Nye Værdier for Transportens Eksterne Omkostninger.”

⁷⁵ Danish Ministry of Transport and COWI, “Rapport om værdisætning af transportens eksterne omkostninger.”



3.2 Brute Force

-ALEXANDER

In order to produce a collision between a known hash value from a known input and any other hash value from the same algorithm (a 2nd pre-image collision), the obvious method for finding such collisions is to keep guessing at random input values in order to encounter one that generates a corresponding output.

The probability of encountering a 2nd pre-image collision for each attempt is independent of the previous attempt and the total probability of encountering a collision with n attempts is thus:

$$P_n(H1_x = H2_x) = 1 - \left(\prod_n \frac{x-1}{x} \right)$$

Where n is the number of guesses.

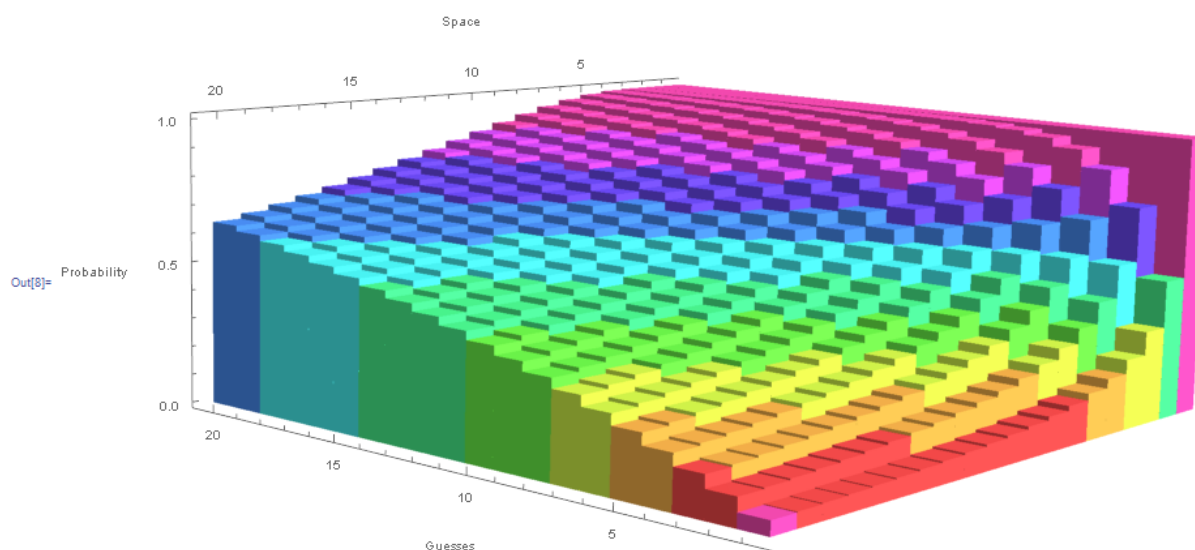


Figure 12: Probability of a generating a 2nd pre-image hash collision as a function of output space size and number of trials. As the size of the space increases, the probability of finding a collision drops and when the amount of guesses increases, so does the chance of finding a collision.

As expected, with an increase in guesses the chance of finding a 2nd pre-image collision increases, however, previous guesses do not contribute to finding new collisions and as the size of the output space increases, the chance of collisions decrease correspondingly.

In modern hash functions with a very large output space, the chance of finding a 2nd pre-image collision just by brute force attempts is a less effective strategy as seen in Figure 13, where a small increase in output size effects a large increase in the amount of guesses needed. Specifically, for each added bit in a binary output, you need twice the amount of guesses in order to have more than 50% chance of finding a 2nd pre-image collision.

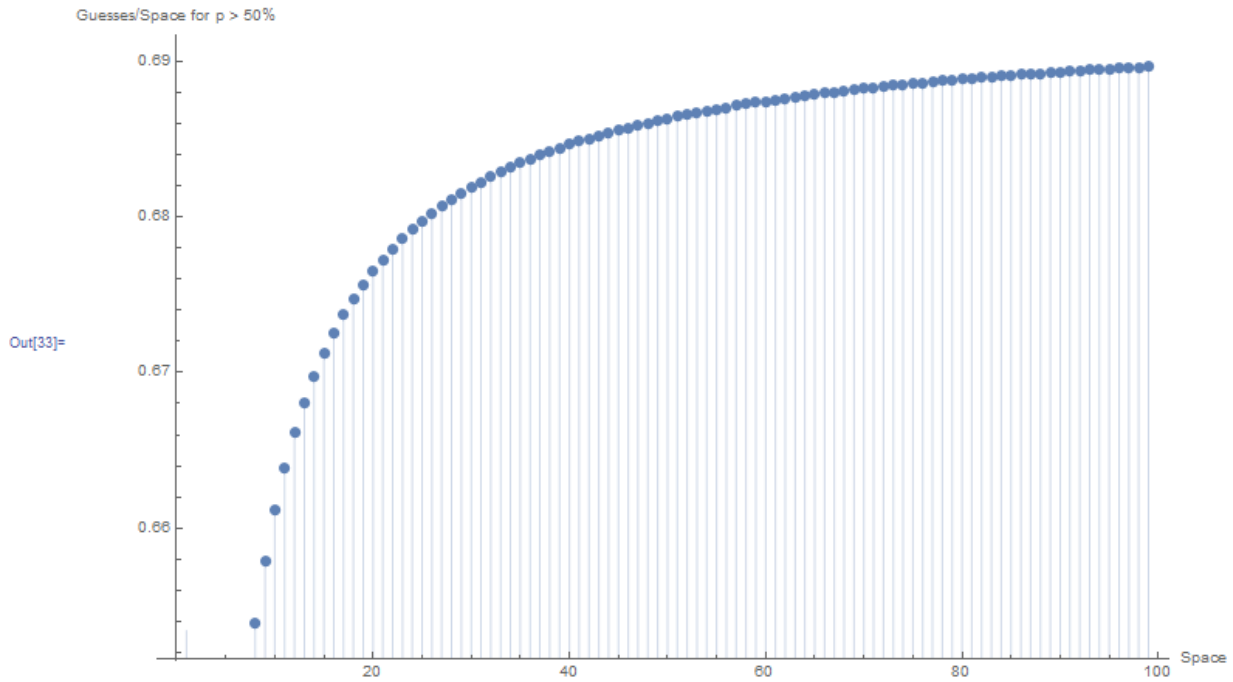


Figure 13: Guesses needed per output space size, as a function of space, for the probability of guessing a 2nd pre-image collision is above 0.5. For anything above trivial sizes it converges to 0.69. This effectively illustrates that for any output space size, the amount of guesses needed would be: 0.69 × the space size.

3.3 Applied Birthday Paradox

-ALEXANDER

Calculating the probability of generating a general hash collision on its own is not very useful beyond proving that a larger output space will reduce the chance of collisions.

However, if the problem definition is relaxed from attempting to produce a hash collision with a specific message (2nd pre-image collision), to producing a hash collision between any two randomly generated messages (general collision), the probability function now takes on another form when every non colliding value generated will be retained for further general collision tests. This phenomenon is also known as the birthday paradox^{77, 78} and the principal equation governing it is as follows:

$$P_n(H1_x = H2_x) = 1 - \left(\prod_n \frac{x - n + 1}{x} \right)$$

Here it can be observed that the general collision chance rapidly grows when the amount of known hashes rises. The cause of this, is that for every newly generated value it has to be tested against all previously generated values and as such the impact of a generated value increases as can be seen by Figure 14 and Figure 15 and Figure 16 below.

⁷⁷ Weisstein, "Birthday Problem."

⁷⁸ "Combinatorics (2.6) The Birthday Problem (2.7) - bday_14-Handout.pdf."

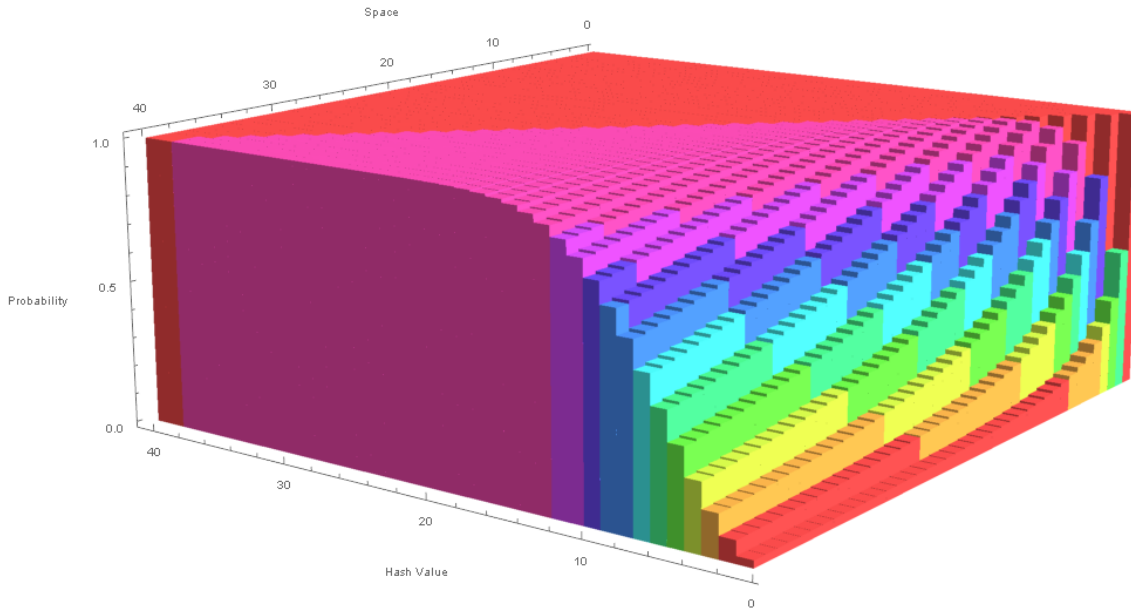


Figure 14 Probability of a collision between two hash values as a function of guesses and the outcome space, when guesses are preserved and added to the test set. The probability of a collision rises with more guesses and falls with a larger output space, however the amount of guesses have a much larger impact than before.

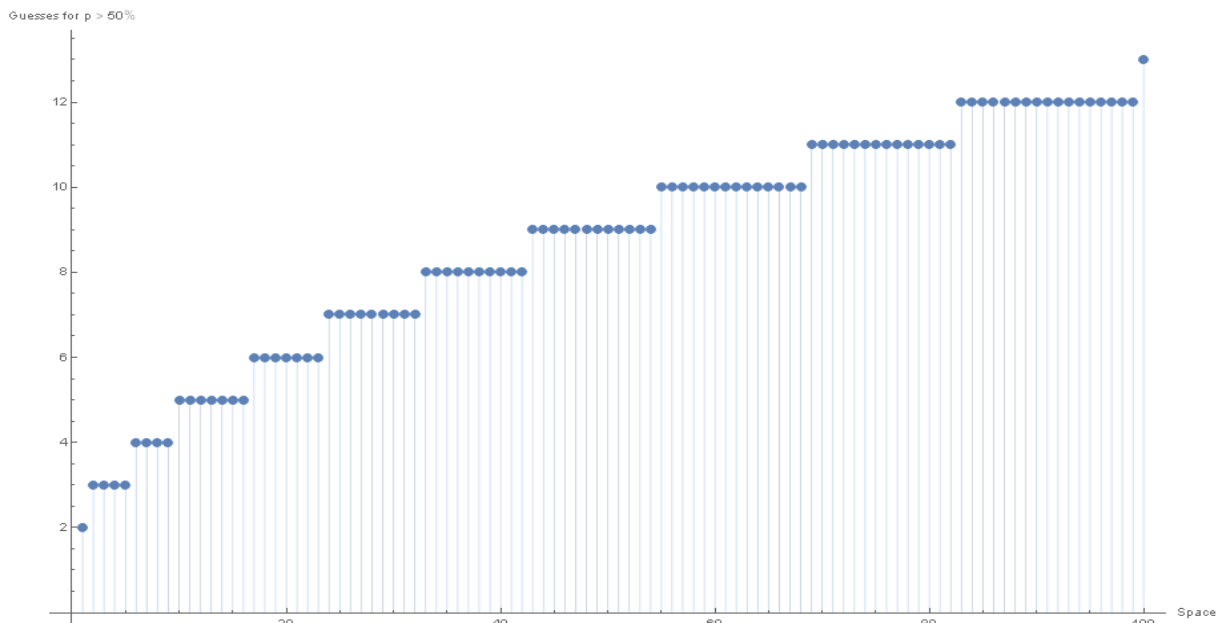


Figure 15 The number of guesses before the probability is above 50%. This time, the amount of guesses needed is not normalized by the output space, leading to an “exponential decay” relationship between the guesses and the probability of success.

As a side note, this probability calculation will take a prohibitively long time to calculate for large values of outcome space, and therefore the Taylor approximation is then found instead such that It can be applied to the spaces of modern hash functions which typically range well above billions of possible outcomes:

$$P_n(H1_x = H2_x) = 1 - e^{\frac{-n(n-1)}{2x}}$$

It can be seen that the error function has a large initial error, however this can safely be ignored since the original function can be used with little effort with low input values.

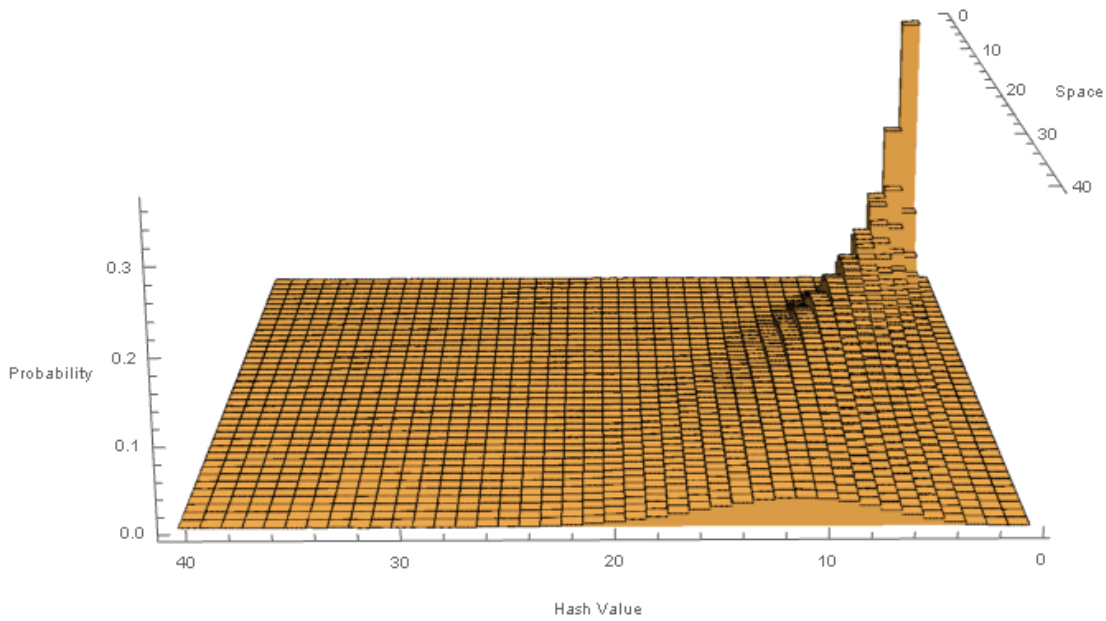


Figure 16 Error function of the Taylor approximation

3.4 Moore's Law

-ALEXANDER

Cryptographic algorithms are primarily replaced when their actual strength proves to be inadequate in the face of an attacker. The two primary reasons for this are cryptanalysis progression, demonstrating weaknesses in algorithms not previously known and the development of ever more powerful hardware.

The best known model for predicting technological advancement is known as Moores law⁷⁹ and it is based on the observation that transistor counts in modern chips doubles approximately every 18 months.

This is due to a variety of factors such as downscaling of transistor size, better packing techniques of chips higher clock frequencies and new technology developments.

This progression is not constant and have been shown to slow down recently, however the claim that computing power roughly doubles each year is still somewhat true.

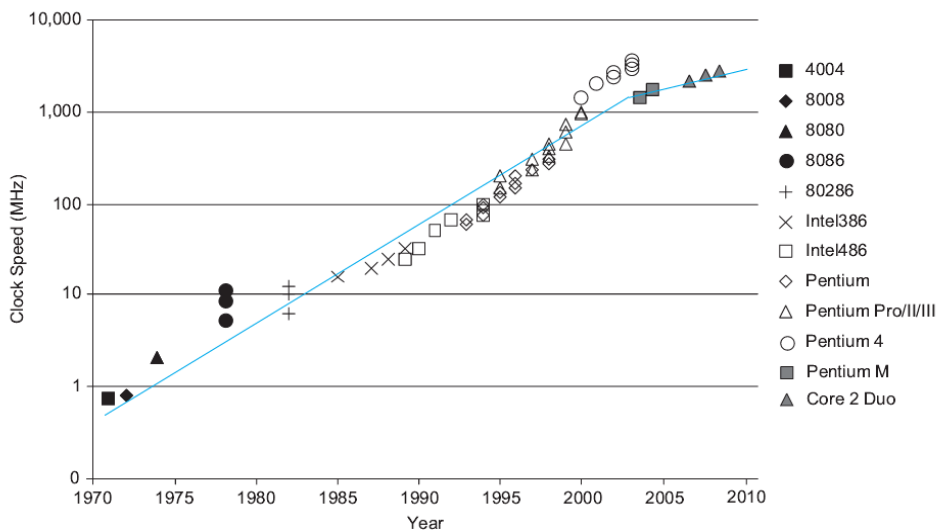


Figure 17 Clock speed of intel processors⁸⁰

⁷⁹ Weste and Harris, *Integrated Circuit Design*.

⁸⁰ Ibid.

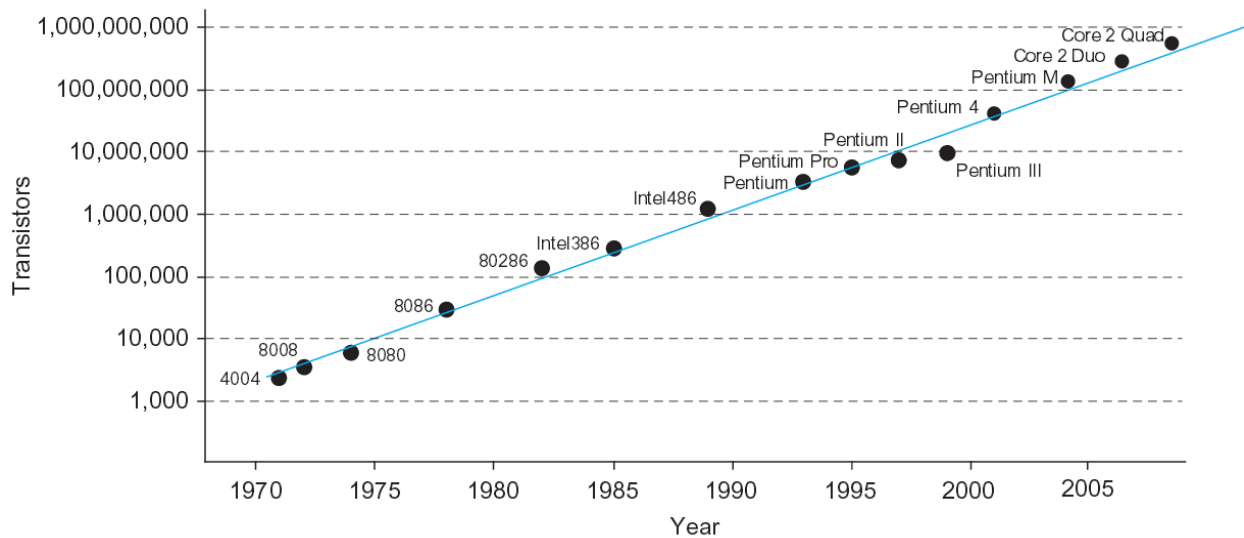


Figure 18 Transistors in Intel Processors⁸¹

4 X.509 Structure

-ALEXANDER

While modern operating systems are typically shipped with tools making them able to interpret or generate an X.509 certificate, being able to modify them is generally not a service offered, since the fact that they are signed by a digital signature precludes this.

This section will focus on describing the structure of a X.509 certificate. It is done by breaking down a concrete example of a certificate (see appendix chapter 20.1 Example Certificate, page 115). A special consideration will be given to structures that needs to be changed in order to produce a valid forged certificate⁸².

4.1 Encapsulation

-ALEXANDER

X.509 certificates is a flexible container that can be encapsulated in several formats based on the type of contents as well as the intended usage: .pem, .cer, .crt and .der are raw certificate containers, with the added ability of having the content base64 encoded. .p7b, p7c .p12 and .pfx are enveloped carriers, allowing more complex meta data to be carried together with the certificate data, which is typically signed content.

While the certificate formats primarily vary by implicit declaration of intended use, they are by design capable of filling overlapping roles. This implies that an interpreter of a certificate need to parse it by content instead of by extension.

Common for all types is that they are encoded in the ASN.1 DER object notation scheme⁸³.

The DER scheme is designed with portability in mind as it encodes location variant data to a binary format which allows for a platform independent unambiguous encoding, with the added extra constraint that any two described objects (e.g. certificates) expressing the same content will look exactly the same in encoded form.

⁸¹ Ibid.

⁸² "RFC 2459 X509 Cert - Obsolete."

⁸³ "DER Encoding of ASN.1 Types (Windows)."



The DER Scheme describes a set of primitives as well as how to construct compound types based on these primitives. It is an implementation of the generic Tag-Length-Value encoding for arbitrary data structures, which are constructed by the primitives: BIT STRING, BOOLEAN, INTEGER, NULL, OBJECT IDENTIFIER, OCTET STRING, BMPString, IA5String, PrintableString, UTF8String, SEQUENCE & SET.

Each type is identified by a specific byte, meaning that the DER format can only specify up to 256 basic types. After the type identifier, the length value designates the length of the following object as an unsigned integer. If the length exceeds what can be described by 7 bits, bit 7 is set and the bits 6 through 0 describes the amount of bytes used to describe the length (which then follows).

An interpreter then reads the described amount of bytes as the length of the value field.

With the exception of the SEQUENCE and SET, the primitives represent a single terminated unit. SEQUENCES and SETS however allows for the construction of more advanced types.

SEQUENCES are used extensively in X.509 and are the foundation for building compound structures or classes. They can contain an arbitrary amount of the ordered primitive types (including other SEQUENCES), which is used in conjunction with an interpretation schema (may be implicit) with the final result being compound objects.

In the case of the X.509 certificate, the top level sequence "Certificate", containing a TBS(to be signed) certificate, a signature algorithm descriptor and the actual signature data.

Only the content of the TBS certificate is hashed and stored in the signature. If padding the content of a certificate, this field needs to be updated.

The TBS Certificate contains the actual certificate data with the following fields:

- Version of the certificate which is an explicit integer either v1(0), v2(1) or v3(2). Only the v3 certificates support arbitrary extensions and can support padding.
 - Serial number of the certificate.
 - Signature field, which must match the outer signature field for consistency.
 - Issuer name, which is a compound type describing the X.509 ASN.1 Distinguished name of the issuer.
 - Validity, Timestamps for the period for which the certificate is valid.
 - Certificate subject. Principal name of the owner of the certificate in the same format as the issuer, aliases can be found in the subject alternate names extension.
 - Subject public key, contains an algorithm id and the key material.
- If the certificate version is above version 1, the following extra fields are present:
- Issuer Unique id / Subject unique id.
 - If the certificate version is above version 2, the extensions field is present which allows for proprietary extensions of the standard. The following standardized extensions are important to remember:
 - o Subject key identifier, a hash of the public contained in the certificate. It is either the **SHA-1** of the subject public key or a marker + the last 60 bits of the hash of the subject public key.
 - Since this field is optional it can be deleted if needed. However, conforming **CA** certificates must include this field.
 - Netscape comment, a freeform comment field which can be used for padding.



4.2 Certificate modification

-ALEXANDER

The goal of understanding the certificate structure is to identify what fields to modify in order to make a fraudulent certificate.

In order to forge an imitating certificate, the only fields to change will be the subject public key (where a new one will be inserted in place of the old), the alternate subject name (in order to make the certificate cover a broader range of domains), the subject key identifier and a custom extension to allow padding will be inserted at the end which will be modified to make the certificate hash to the same value as the original it was based upon. A good candidate for padding would be the Netscape comment extension since it allows freeform text content, while being largely ignored for most applications⁸⁴.

A choice has to be made when determining whether to modify the subject key identifier, as it does not serve a cryptographic purpose, but is used for quickly identifying a certificate. By modifying this field, the certificate loses the ability to take the place of the original certificate in indexing and chain-building operations, but a more thorough verification would mark the certificate as fraudulent since the identifier and the public key it is based upon do not match.

In the example, it is decided to let it stay unmodified as it allows for a quick rejection of the fraudulent certificate in thorough certificate checking utilities and as such it limits the potential damage the certificate could cause should it be compromised.

It is important to remember to align the freeform padding such that it maximizes the available freeform input space, without exceeding the input restrictions of **SHA-1**. The goal here is to make sure that when brute-forcing the result, the algorithm only needs to process one 512bit chunk of data, instead of multiples.

The concrete modified certificate along with the certificate it is based on can be seen in appendix chapter 20.1.1 Modified certificate overview, page 115.

5 Current use of SHA-1

-LARS

Hashing is a fundamental part of IT security hence **SHA-1** has many uses, falling into all three of the core security objectives defined by, among others, Octave Allegro⁸⁵ to be:

Confidentiality, Integrity, and Availability

Masquerading other information to look like an asset owned by others fall outside these categories and how they are described in the Octave Allegro worksheet seen below:

		(4) Outcome	<input type="checkbox"/> Disclosure	<input type="checkbox"/> Destruction
		What would be the resulting effect on the information asset?	<input type="checkbox"/> Modification	<input type="checkbox"/> Interruption

Figure 19 Worksheet assisting an Octave Allegro assessment, worksheet 10

⁸⁴ Stevens et al., "Short Chosen-Prefix Collisions for MD5 and the Creation of a Rogue CA Certificate."

⁸⁵ Caralli et al., "The OCTAVE Allegro Guidebook, v1. 0."



Masquerading is not Disclosure, nor Destruction, Modification or Interruption, hence looking through the applications of hashing another group emerge:

Authentication / non-repudiation^{86, 87, 88}, that also applies to hash usage.

The following subchapters will show some examples accompanied with a shortlist of security issues.

5.1 Code Signing

-LARS

Verifying code to be from a specific source does not only require Integrity, but also uniquely identify the sender: Non-repudiation.

Using **RSA** is a common way to do this, with a publicly available signature and a key of several thousand bits. The part being cryptographically signed is only a digest of the full code though, making the digest algorithm a viable attackvector when it poses less difficulty than the signing algorithm and key⁸⁹.

Given that the key need to be kept secret, there are few ways of handling that:

Deletion, **H**ardware **S**ecured **M**odule(**HSM**), Shamir Secret Sharing and splitting it, which has been explored in a later chapter:15.1.1 Shamir Secret Sharing, page 95.

5.2 Document Signing

-LARS

In the wake of the green digitization wave, signatures have moved into the digital realm^{90, 91}, even with European directives equating them to physical ones when a qualified certificate is used⁹². With the **P**ortable **D**ocument **F**ormat(**PDF**) being specified in **ISO** 32000 January 2008, so was the signature dictionary listing the 3 mandatory SubFilters as: `adbe.x509.rsa_sha1`, `adbe.pkcs7.detached`, and `adbe.pkcs7.sha1`⁹³(PAGE 467). With X.509 and pkcs implementations in a wide array of software packages they are not always implemented according to specifications in several cases leading to vulnerabilities^{94, 95}.

5.3 BitTorrent Protocol

-ALEXANDER

The **B**it**T**orrent protocol⁹⁶, is one of the most widely used peer-to-peer file sharing mechanisms in general use today⁹⁷. It relies on a centralized index, called a tracker, which uses **SHA-1** hashes to organize the traffic between participating peers wishing to download a resource. It also uses **SHA-1** to verify file integrity.

⁸⁶ Gürgens and Rudolph, "Security Analysis of Efficient (Un-) Fair Non-Repudiation Protocols."

⁸⁷ Barker, "Recommendation for Key Management: Part 1: General (Revision 4) DRAFT SP800-57," 57.

⁸⁸ "Core PKI Services: Authentication, Integrity, and Confidentiality."

⁸⁹ Stevens et al., "Short Chosen-Prefix Collisions for MD5 and the Creation of a Rogue CA Certificate."

⁹⁰ Merkle, "A Certified Digital Signature."

⁹¹ Bellare and Miner, "A Forward-Secure Digital Signature Scheme."

⁹² *Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on a Community Framework for Electronic Signatures.*

⁹³ ISO, "ISO 32000-1."

⁹⁴ Park et al., "Security Analysis on Digital Signature Function Implemented in PDF Software."

⁹⁵ Itoh et al., "Forgery Attacks on Time-Stamp, Signed PDF and X.509 Certificate."

⁹⁶ Cohen, "The BitTorrent Protocol Specification."

⁹⁷ Sandvine, "Sandvine Global Internet Phenomena Report - 2H 2014 - 2h-2014-Global-Internet-Phenomena-Report."

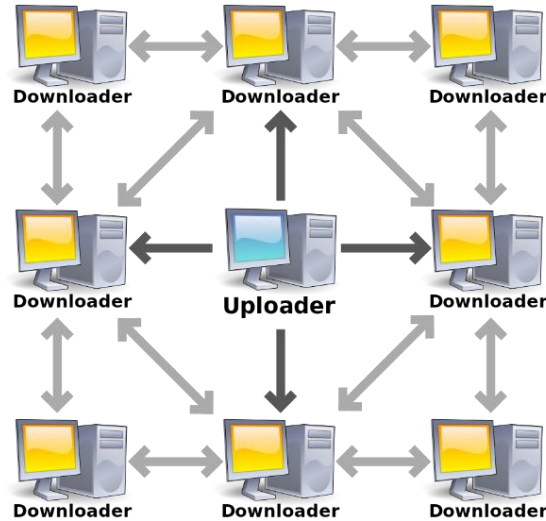


Figure 20: Classic BT architecture, where downloaders accelerate the collective download speed by uploading their data to peers. By ⁹⁸

Compared to a traditional direct HTTP download, the process of downloading a file through a torrent network is a bit more complicated, as it involves setting up a dedicated tracker, constructing a .torrent metadata file, serving the metadata file through a webserver and hosting the data through a torrent client.

Fortunately, a tracker can host multiple torrent swarms (a torrent swarm is the term for an amount of torrent data with a group of interested peers), and the most common usage scenario is to produce the .torrent file yourself and use a publicly available tracker for distribution, such as ThePirateBay, Demonoid etc. ^{99, 100, 101}

5.3.1 BitTorrent Metadata Files

-ALEXANDER

At the heart of a torrent swarm is the torrent metadata file, the purpose of which is to describe, the structure of the data to be shared. It is encoded in the Bencode format as described in ¹⁰², and every peer connecting to a swarm will possess this torrent file.

5.3.2 Structure

-ALEXANDER

At the top level, the meta info file is a dictionary with the two fields: *info* and *announce*. ¹⁰³(p. metainfo files) The *announce* field is the **URL** of the tracker hosting the content and the *info* is a dictionary containing further information regarding the data. Note that multiple trackers can be specified for resiliency purposes.

⁹⁸ Martin, *BitTorrent Network*.

⁹⁹ "Top 10 Most Popular Torrent Sites of 2015."

¹⁰⁰ "Top 10 Most Popular Torrent Sites of 2014."

¹⁰¹ "The 5 Most Popular BitTorrent Trackers."

¹⁰² Cohen, "The BitTorrent Protocol Specification."

¹⁰³ Ibid.



5.3.2.1 Info

-ALEXANDER

The *info* dictionary contains the three mandatory keys: *name*, *piece length* and *pieces* as well as exactly one of the two optional keys: *length* (in the case of a single file) or *files* (in the case of a multi-file torrent).

Of particular interest is the *piece length* and *pieces entries*, as these contain hash data:

All the data in a complete **BT** download, independent of file structure are split into a number of pieces, where each piece is “piece length” bytes long (except the last piece which may be shorter and implicitly zero padded). The **SHA-1** hash of all pieces are concatenated in order of appearance and set as the *pieces* key. ¹⁰⁴

The *piece length* and *pieces* entries thus constitute the integrity validation mechanism of a torrent download.

The **BT** metainfo protocol is standardized by the **BT** Community Forum, which coordinate all of its development as well as monitor and oversee what extensions needs further work¹⁰⁵.

5.3.2.2 Typical BitTorrent File

-ALEXANDER

While there is a great deal of flexibility when constructing a torrent file for a set of data, a set of guidelines exist which users are encouraged to use when sharing content, especially regarding piece sizes, where the official standard strongly recommends piece lengths be a power of two and if possible, $2^{18} = 256K$ bits specifically (older revisions recommend $2^{20} = 1M$ bits). In the case of very large files, it is recommended to choose another power of two which make the total amount of pieces between approximately 1000-2000. ¹⁰⁶

These sizes are recommended because, pieces are typically fetched from a single peer and very small or very large pieces either spend a rather large amount of time setting up the connection compared to the amount of data transferred or risk turning the torrent protocol into a single direct transfer due to errors respectively.

5.3.3 Tracker protocol

-ALEXANDER

A peer periodically communicates with the trackers attached to a swarm in order to update its list of potential peers as well as report its own metrics, such as uploaded / downloaded amount (this allows a tracker to optimize swarm routing). ¹⁰⁷(p. Trackers)

All communication is done with **HTTP** GET requests and the tracker responds with Bencoded dictionaries.

The only thing of particular note in this protocol is the handshake phase, where the torrent client informs what port/**IP** it expects future users to contact it with.

The tracker is also responsible for monitoring the general Health of a swarm, here the Health specifically refers to the availability of a particular torrent, as it is entirely possible for parts of the data to be completely unavailable (a seeder may have gone offline). This is typically measured in percentage of the complete data available, where a Health below 100% effectively means that it is not possible to download the complete set. An example of an availability statistic can be seen in Figure 21.

¹⁰⁴ 3rd and Jones, “RFC3174 - US Secure Hash Algorithm 1 (SHA1).”

¹⁰⁵ Harrison, “Index of BitTorrent Enhancement Proposals.”

¹⁰⁶ Vuze Team, “Vuze Open-Source BitTorrent Client Documentation.”

¹⁰⁷ Cohen, “The BitTorrent Protocol Specification.”

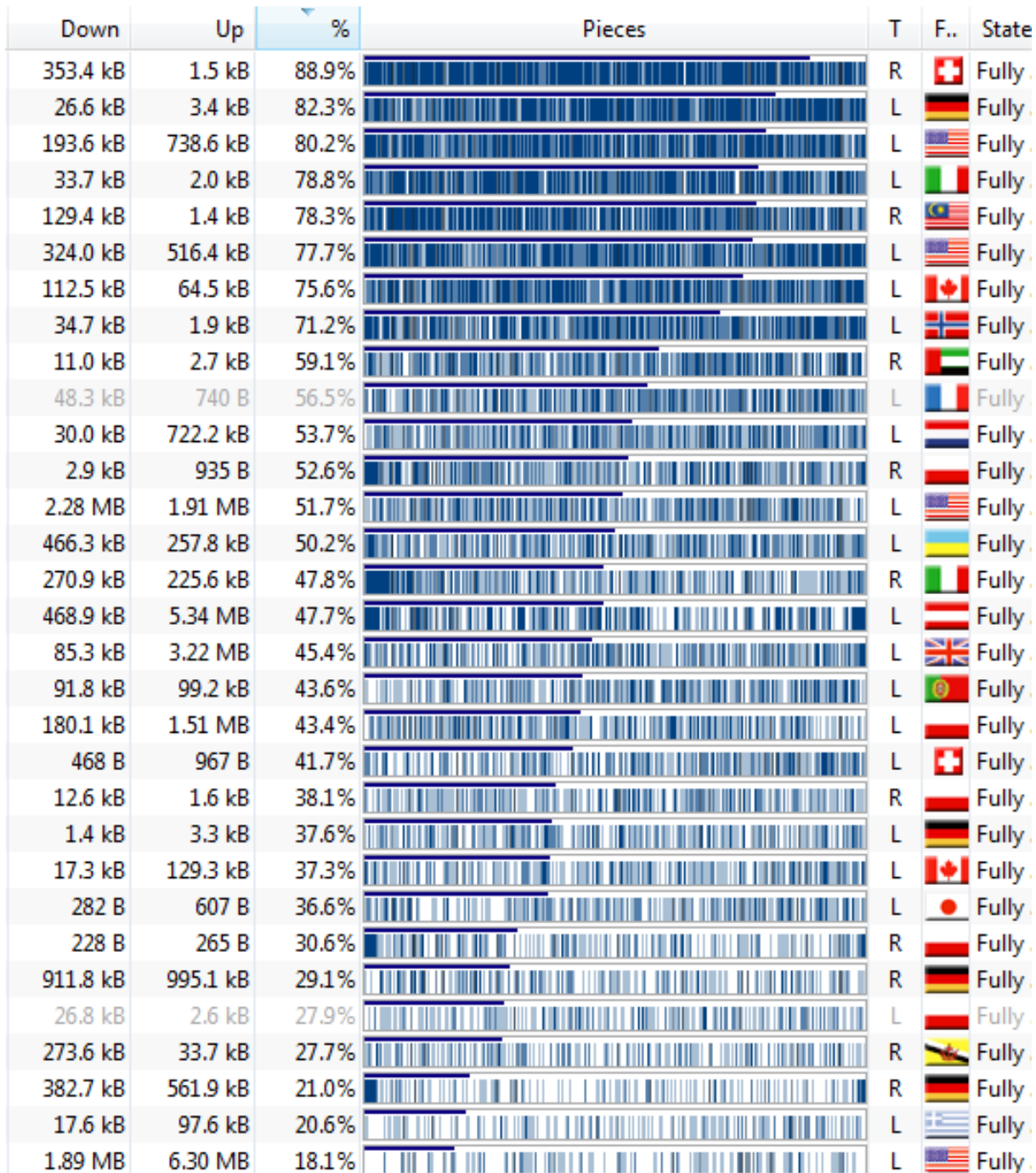


Figure 21: Availability map of an example torrent distribution in a swarm. Notice overlapping coverage areas, this is to be considered a healthy swarm since the availability is high.

By ¹⁰⁸

¹⁰⁸ Vuze Team, *Peers with Pieces*.



5.3.4 Peer Protocol

-ALEXANDER

BT clients (Peers) communicate between each other using a peer protocol based on **TCP** (or in specific cases on a proprietary version of **TCP** called uTP¹⁰⁹). It serves the purposes of both controlling the internal flow of data and state. ¹¹⁰(p. Peer protocol)

Connections between peers start out as *choked* and *not interested*, where *choked* express “The local client will not send to the remote client” and *interested* express “The sender is interested in data the recipient possesses”. Connections are established by a **TCP** handshake followed by a header consisting of the decimal length prefixed string: “BitTorrent Protocol” and some data identifying the current download.

Each individual client can decide whom to *unchoke* based on its own algorithms, but it should always update all peers on its *interested* status.

The list of known peers is periodically updated by contacting the tracker(s) associated with the swarm.

There are many implementation specific choices for transferring a complete torrent using this protocol, but in general, pieces are requested at random from random sources, with individual blocks in a piece from the same source.

5.3.5 DHT

-ALEXANDER

In an effort to reduce the reliance on central trackers, modern implementations **BT** protocol utilize **DHTs**.

DHT, or a Distributed Hash table, is an implementation of a traditional Hash table spread across several individual connected nodes, with the explicit goal of scaling well to extremely large datasets. ¹¹¹(C. 1)

As a concept, there is no specific standard implementation of a **DHT**, but it typically consists of two primary components: The Keyspace partitioning algorithm and the Overlay Network topology and routing mechanisms. ¹¹²(C. 2)

Combined, it should allow a client to query a network with a hash value and receive its corresponding stored data, just like a normal Hash table.

5.3.5.1 Usage in BitTorrenting

-ALEXANDER

A specific type of **DHT** is employed in order to allow a peer to download torrent data without connection to a tracker. Based on the *Kademlia* **DHT** algorithm, every node stores routing data as well as key data. ¹¹³(3) ¹¹⁴

When clients join the *Kademlia* **DHT** for the first time, they generate a 160 Bit random ID which it keeps permanently. This ID is what peers use to calculate their mutual distances as well as the distance to a specific info-hash (Also a 160 Bit value, which is the output of a **SHA-1** operation).

To calculate the distance between two nodes, or a node and an info hash, the two values are XOR’ed together and the result is interpreted as an unsigned integer. This measure does however not have any relation to

¹⁰⁹ Nordber, “uTorrent Transport Protocol.”

¹¹⁰ Cohen, “The BitTorrent Protocol Specification.”

¹¹¹ Zhang et al., *Distributed Hash Table*.

¹¹² Ibid.

¹¹³ Grunthal, “Efficient Indexing of the BitTorrent Distributed Hash Table.”

¹¹⁴ Loewenstern and Nordberg, “DHT Protocol.”



physical distance or connectivity, but it does provide an easy to calculate measure, that will never change, no matter how the underlying topology is constructed.

Each peer maintains a list of known “close” good nodes, based on their performance and it is the responsibility of a peer to keep an up-to-date routing table. When the peer wishes to fetch torrent metadata, it will query the nodes closest to the data (again simply attained by XOR’ing the hash and the peer id of the neighbors) and they will either respond with the torrent metadata with a list of peers that are closer to the metadata. This way a peer will traverse the **DHT** swarm until it reaches the data.

Beyond being able to fetch metadata from the **DHT** swarm, nodes act as trackers for info hashes which are sufficiently close to them. (this topic, as well as error correction and swarm maintenance is beyond the scope of this project.)

It is important to note that the **DHT** offers no guarantee to return the complete set of peers, as the swarm can easily fragment if not all tracking peers of a torrent is equally close to the new peer that wishes to join.

Also if a creator of a torrent explicitly only wishes to use **DHT**, the standard allows for a Magnet link to embed node ID’s of tracker clients instead of tracker **URL**’s directly.

5.3.6 PEX

-ALEXANDER

PEX, or Peer **EX**change, is the umbrella term for a set of protocols designed to let peers discover more peers in a swarm. There are multiple distinct and incompatible protocol versions, but they achieve the same result and most modern torrent clients support many, if not all major, versions. ¹¹⁵

Common for all versions is, that a conforming client will periodically (max once pr. minute) inform other members of the swarm who have joined and left the swarm since the last update. Whether this is done by push or pull mechanics is implementation specific, based on the carrying layer (either as messaging protocol or the mainline extension protocol).

When this technique is combined with **DHT** tracking, it reduces the potential for swarms to segment which **DHT** normally are subject to and it allows the tracking peers to self-coordinate, which drastically improves performance and coverage.

5.4 Content Distribution Networks

-ALEXANDER

The **BitTorrent** protocol also sees use as an imbedded data transfer protocol, for instance in conjunction with content delivery networks¹¹⁶, where large amounts of data needs to be mirrored across multiple nodes with a significant physical distance between them. As long as there is not a requirement for high throughput between two individual nodes or strict real time requirements, **BitTorrent** is extensible enough to facilitate these custom domains.

5.5 openPGP

-LARS

openPGP (**open** **P**retty **G**ood **P**rivacy) is a Public-Private key encryption standard intended to secure e-mail communication and data where hashing is used to ensure integrity as well as authentication. But because everyone is able to upload to a global registry of keys and assigning any name to the keys impersonation is

¹¹⁵ Theory Team, “BitTorrentPeerExchangeConventions - Theory.org Wiki”; Vuze Team, “Peer Exchange - VuzeWiki.”

¹¹⁶ “8 Legal Uses For BitTorrent.”



possible hence it does not have non-repudiation. It is an open source fork of the now commercial PGP and is governed by RFC4880¹¹⁷, where SECTION 9.4 details SHA-1 as the only mandatory hash function openPGP software is required to support, with a list “[1]MD5,[2]SHA-1,[3]RIPE-MD/160,[4-7]Reserved,[8]SHA256,[9]SHA384,[10]SHA512,[11]SHA224” and the notion that MD5 is deprecated, but without such a notice on SHA-1.¹¹⁸

Furthermore, RFC4880 SECTION 13.3.2. details, that for practical reasons a sender can specify the hashing algorithm they want the recipient to use for replies e.g. an older weaker hashing algorithm.

This opens up for a downgrade attack vector weakening the security to at least SHA-1.

With SHA-1 being the mandatory default, SHA-1 is currently the fall-back if nothing is specified leading to most software not specifying a hashing algorithm.

“Since SHA1 is the MUST-implement hash algorithm, if it is not explicitly in the list, it is tacitly at the end. However, it is good form to place it there explicitly.”

-RFC4880 SECTION 13.3.2.

We recommended a revision of the RFC and implementation to include the request for stronger hashing algorithms.

As public keys can be appended with the information of preferred hashing algorithm openPGP public key servers can be used to inform contacts of a preference on the use of a stronger hashing algorithm.

A study of the most popular key servers (pool.sks-keyservers.net, keys.gnupg.net, pgp.mit.edu [popular, but not recommended]) is recommended for future work, using the resources at

<http://www.staff.science.uu.nl/~penni101/wotsap/> and <http://pgp.cs.uu.nl/archive/>.

5.6 Law

-LARS

In 1999 the European Union Directive on “a Community framework for electronic signatures” was made¹¹⁹, ratified into Danish law the next year¹²⁰, which set a framework with guidance on certificate and key management.

Following are some quotes from the directive:

(10) The internal market enables certification-service-providers to develop their cross-border activities with a view to increasing their competitiveness, and thus to offer consumers and businesses new opportunities to exchange information and trade electronically in a secure way, regardless of frontiers; in order to stimulate the Community-wide provision of certification services over open networks, certification-service-providers should be free to provide their services without prior authorisation; prior authorisation means not only any permission whereby the certification-service-provider concerned has to obtain a decision by national authorities before being allowed to provide its certification services, but also any other measures having the same effect;

(18) The storage and copying of signature-creation data could cause a threat to the legal validity of electronic signatures;

Danish law text on this:

§ 10. Et nøglecenter skal registrere og opbevare alle relevante oplysninger om certifikaterne i en rimelig periode, dog mindst seks år.

¹¹⁷ Shaw et al., “OpenPGP MessageFormat - RFC 4880.”

¹¹⁸ Ibid.

¹¹⁹ Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on a Community Framework for Electronic Signatures.

¹²⁰ Lov Om Elektroniske Signaturer (Act No. 417 of 31 May 2000 on Electronic Signatures).



Stk. 3. Nøglecentre må ikke opbevare eller kopiere de personers signaturgenereringsdata, som nøglecentret gennem udstedelsen af certifikater måtte have fået kendskab til.
 -LOV OM ELEKTRONISKE SIGNATURER (ACT NO. 417 OF 31 MAY 2000 ON ELECTRONIC SIGNATURES)¹²¹

(20) Harmonised criteria relating to the legal effects of electronic signatures will preserve a coherent legal framework across the Community; national law lays down different requirements for the legal validity of hand-written signatures; whereas certificates can be used to confirm the identity of a person signing electronically; advanced electronic signatures based on qualified certificates aim at a higher level of security; advanced electronic signatures which are based on a qualified certificate and which are created by a secure-signature-creation device can be regarded as legally equivalent to hand-written signatures only if the requirements for hand-written signatures are fulfilled;

Summarising: it should be free(libre) for everyone to make a certificate service, private keys should be kept only by the user to ensure non-repudiation, digital signatures living up to the directive requirements have same legal binding as a hand-written signature.

5.7 Summary of KPI

-LARS

The security **KPI** of the hashing applications mentioned in this sub chapter are listed in the table below for an easy overview:

Table 2 KPI for SHA-1 use in different domains

	<i>Confidentiality</i>	<i>Integrity</i>	<i>Availability</i>	<i>Authentication / Non-repudiation</i>
Code signing	YES	YES		YES
Document signing	YES	YES		YES
P2P		YES	YES	
CDN		YES		
Law	YES	YES		YES

PART 2 HAZARD IDENTIFICATION

-LARS

In the following chapters several hazards are described.

Hazard identification provide a foundation for Part 3 that describe the methods to ascertain the probability of a **SHA-1** general and 2nd pre-image collisions, and along with the hazards identified in this part and the worst case scenarios form a base for a risk assessment.

¹²¹ Ibid.



6 Hazard Identification

-LARS

As hashing is a fundamental part of IT security hence **SHA-1** has many uses, falling into all three of the core security objectives defined by among others Octave Allegro¹²² to be:

Confidentiality, Integrity, and Availability

Looking through the applications of hashing another group emerge:

Authentication / non-repudiation^{123, 124}.

The following subchapters will show some examples accompanied with a shortlist of security issues.

6.1 Apple Update Distribution

-LARS

The story in IT circles goes like this:

“On every major OSX and IOS update Apple makes a new certificate, sign the OS and discard/delete the key (rather than saving it in a HSM) so no one else can ever get hold of it and sign malicious data.”

Sadly, it has been impossible to verify it, neither through documentation, nor E-mail or telephone contact with Apple.

It has however been possible to find the number of root certificates in that use **SHA-1** in OS X (217)¹²⁵ and IOS 5 & 6 (155)¹²⁶. As well as the guide to “*verify the authenticity of manually downloaded Apple software updates*”¹²⁷ using **SHA-1** as well.

Causing the Apple update service (at least the manual ones) and 100+ root certificates to be vulnerable to a **SHA-1** attack.

6.2 Document Signing

-LARS

Signing documents is done in the same way as code, but has a deeper legal impact as digitization has moved previously handwritten signatures into the digital realm^{128, 129}. With the European directives equating digital signatures to physical ones when a good enough qualified certificate is used¹³⁰. As **PDF** was specified in **ISO 32000** January 2008, so was the signature dictionary listing the 3 mandatory SubFilters as:

`adbe.x509.rsa_sha1`, `adbe.pkcs7.detached`, and `adbe.pkcs7.sha1`¹³¹ (PAGE 467).

With `x509` and `pkcs` implementations in a wide array of software packages they are not always implemented according to specifications in several cases leading to vulnerabilities^{132, 133}.

¹²² Caralli et al., “The OCTAVE Allegro Guidebook, v1. 0.”

¹²³ Gürgens and Rudolph, “Security Analysis of Efficient (Un-) Fair Non-Repudiation Protocols.”

¹²⁴ Barker, “Recommendation for Key Management: Part 1: General (Revision 4) DRAFT SP800-57,” 57.

¹²⁵ “Lists of Available Trusted Root Certificates in OS X - Apple Support.”

¹²⁶ “iOS 5 and iOS 6: List of Available Trusted Root Certificates - Apple Support.”

¹²⁷ “How to Verify the Authenticity of Manually Downloaded Apple Software Updates - Apple Support.”

¹²⁸ Merkle, “A Certified Digital Signature.”

¹²⁹ Bellare and Miner, “A Forward-Secure Digital Signature Scheme.”

¹³⁰ Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on a Community Framework for Electronic Signatures.

¹³¹ ISO, “ISO 32000-1.”

¹³² Park et al., “Security Analysis on Digital Signature Function Implemented in PDF Software.”

¹³³ Itoh et al., “Forgery Attacks on Time-Stamp, Signed PDF and X.509 Certificate.”



Even with a correct implementation electronic document signatures relying on **SHA-1** based **X.509** certificates are at risk.

Leading **SHA-1** based **X.509** certificates to be a prime candidate for parallel processing to find a 2nd pre-image collision for the following reasons:

1. Due to the Merkle-Damgård construction¹³⁴ all but the last rotation set can be pre-computed
2. Revocation, meaning less damage if a collision is found
3. No need to generate and save random input data
4. Verified by a government **TSL**
5. High value target
6. Recently in use

The candidate chosen for collision testing was therefore the European Commission **X.509** certificate valid March 2013 to March 2015.

As verified by the **TSL** of Bulgaria¹³⁵, Cyprus¹³⁶, Denmark¹³⁷, France¹³⁸, Iceland¹³⁹, Italy¹⁴⁰, Latvia¹⁴¹, Luxembourg¹⁴², Malta¹⁴³, Poland¹⁴⁴, Romania¹⁴⁵, United Kingdom¹⁴⁶.

With the Subject Key Identifier BF:85:2C:A8:B6:B5:1C:ED:3E:FB:16:BF:02:51:10:BO:90:79:71:F3 as well as the whole **ASN.1** encoded certificate.

6.3 Certificates

-LARS

While current European Commission certificate only allows for authentication, specific to ec.europa.eu, the possibilities extend to also include money transfers, server authentication and other domains with a forged certificate.

With the law equating digital and physical signatures, as described in chapter 5.6 Law, page 41.

Forged certificates makes is possible to take up loans, transfer ownership of property, cars, certify university records and any other task requiring a signature.

With the exception that a digital signature can be proven to belong to and represent an organisation like a company or European Commission with a set amount of authority as specified in the certificate¹⁴⁷(**CHAPTER 4.2.1.3 KEY USAGE**).

It is possible to do this using a website certificate as a base and then change it to be valid for signing documents and money transfers, rather than just a secure website connection, while still having a certificate chain validating it. It should be noted that the other certificate in the Danish **TSL** uses **SHA-256**¹⁴⁸.

¹³⁴ Stafford E., *On the Design of S-Boxes - Advances in Cryptology*.

¹³⁵ "БЪЛГАРИЯ (BULGARIA) : Trusted List."

¹³⁶ "ΚΥΠΡΟΣ/KIBIS (CYPRUS) : Trusted List."

¹³⁷ "DANMARK (DENMARK) : Trusted List."

¹³⁸ "FRANCE (FRANCE) : Trusted List."

¹³⁹ "ÍSLAND (ICELAND)."

¹⁴⁰ "ITALIA (ITALY) : Trusted List."

¹⁴¹ "LATVIJA (LATVIA) : Trusted List."

¹⁴² "Luxembourg (Luxembourg): Trusted List."

¹⁴³ "MALTA (MALTA) : Trusted List."

¹⁴⁴ "POLSKA (POLAND) : Trusted List."

¹⁴⁵ "ROMÂNIA (ROMANIA) : Trusted List."

¹⁴⁶ "UNITED KINGDOM (UNITED KINGDOM) : Trusted List."

¹⁴⁷ The Internet Society, "RFC 3280 - Internet X.509 Public Key Infrastructure."

¹⁴⁸ "DANMARK (DENMARK) : Trusted List."



6.3.1 Trust 2408

-LARS

While the Danish **TLS**'s second entry, apart from the European Commission says TRUST2408, it is the Norwegian "Nassa Midco AS".

Put in bullet points the ownership chain goes like this:

- Nassa Midco AS (Norwegian Organization number: 913 111 990)¹⁴⁹
- (that prior to march 2014 had the name STARTUP 629 14 AS)¹⁵⁰
- that 100% owns NASSA A/S (CVR 34903360)¹⁵¹
- that 100% owns NETS HOLDING A/S (CVR 27225993)¹⁵²
- that 100% owns NETS A/S (CVR 20016175)¹⁵³
- that 100% owns NETS DANID A/S (CVR 30808460)¹⁵⁴
- that has the secondary name TRUST2408 A/S
- that runs the service NemID & is featured on the Danish **TSL**

Just to be transparent, as the end-user is shown the following when installing the NemID E-mail software:

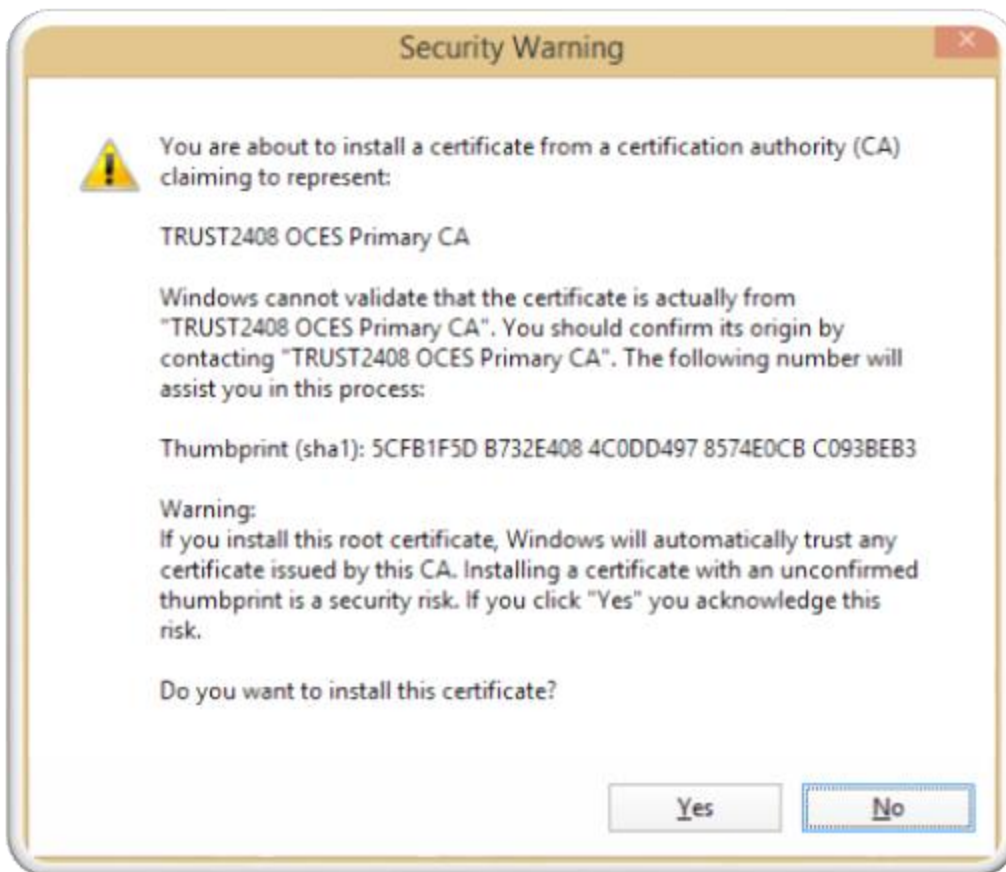


Figure 22 User experience for "Secure mail" - showing the name of TRUST2408 OCES

¹⁴⁹ The Brønnøysund Register Centre (Norwegian Business Registry), "NASSA MIDCO AS Organization Number: 913 111 990."

¹⁵⁰ The Brønnøysund Register Centre (Norwegian Business Registry), "Change of Business Enterprise Name STARTUP 629 14 AS to NASSA MIDCO AS."

¹⁵¹ Virk (Danish Business Registry), "NASSA A/S (CVR 34903360)."

¹⁵² Virk (Danish Business Registry), "NETS HOLDING A/S (CVR 27225993)."

¹⁵³ Virk (Danish Business Registry), "NETS A/S (CVR 20016175)."

¹⁵⁴ Virk (Danish Business Registry), "NETS DANID A/S (CVR 30808460)."



6.4 BitTorrent

-ALEXANDER

A number of attacks against the **BT** protocol already exist, typically mounted in an effort to disrupt or slow down the overall performance of a swarm. These attacks are typically mounted by a copyright holding entity^{155 156}, in response to newly created torrents containing copyrighted material.

The types of known attacks can be broadly categorized as either fake-block attacks or uncooperative-peer attacks.

In both types of attacks, one or more malicious peers advertise that they possess every chunk in the swarm and are willing to share it, essentially acting as a seeder. It should also be noted that they require a large amount of peers to be effective, and modern peer clients implement tools to counter the attacks to an extent.

The most commonly used countermeasure is community distributed blocklists as well as automatic **IP** banning mechanisms.

Blocklists, are files distributed in a **BT** community, with **IPs** of known attackers. Once an **IP** is on the list it can effectively no longer utilize the **BT** protocol in that community. Since **IP** addresses have the potential to be transient, this can lead to unintentional blocking of legitimate peers, which can be hard to reverse.

Automatic **IP** banning works by automatically blocking individual peers in a client, any time they exhibit abnormal behavior (precisely what constitutes abnormal behavior can be client dependent). This will work for normal amounts of malicious peers, but it prevents repeated attacks from the same clients and does not do anything against attackers with rapidly changing **IP** addresses^{157 158}.

6.4.1 Fake-block Attack

-ALEXANDER

In the case of the fake-block attack, the attacker will, upon request of a block, send a block of null/garbage data, and the peer receiving the block will only be able to detect the garbage data when all blocks in the piece are received, essentially corrupting an entire piece with only a small amount of data¹⁵⁹. While an effective **DOS** attack initially, it is quickly stopped by **IP** blacklists.

6.4.2 Uncooperative-peer Attack

-ALEXANDER

The uncooperative-peer attack uses another strategy, whereby it will advertise itself as being in possession of many or all of the pieces of a torrent file, but will silently ignore any incoming requests of blocks, and thus waste time and bandwidth of legitimate users. The attack is further enhanced by, immediately upon retrieval of the request message, the peer will retransmit a handshake and bitmap message, such that the connection is reset instead of dropped, allowing the attack to potentially repeat itself.

A variation of the attack has been observed, using the enormous **IPv6** address space to further improve its effectiveness. It works by simply having access to a practically endless supply of **IP** addresses, effectively making automatic uncooperative peer detection useless, as this mechanism typically operates on an individual address basis which has the same weaknesses as **IP** blacklisting in general.

¹⁵⁵ Torkington, "HBO Attacking BitTorrent - O'Reilly Radar."

¹⁵⁶ Svensson, "Consumer Groups Ask FCC to Fine Comcast."

¹⁵⁷ Dhungel et al., "A Measurement Study of Attacks on BitTorrent Leechers."

¹⁵⁸ Kong et al., "A Study of Pollution on BitTorrent."

¹⁵⁹ Jie Kong and others, 'A Study of Pollution on BitTorrent', in *Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on (IEEE, 2010)*, III, 118–22

<http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5452055> [accessed 25 June 2015].



6.4.3 Leeching

-ALEXANDER

Leeching (only downloading, not uploading) negatively impacts the health of a swarm, since leeching does not contribute to the positive externality by increasing the availability of data, as uploading would. This can theoretically degenerate the protocol performance to that of a regular direct file transfer.

A special case of **BT** use exists, wherein the natural throttling of leeching is circumvented by uploading fake blocks. This is a variation of the fake block attack, however with a different goal in mind.

The rationale is that leeching downloads can temporarily boost the local upload score and thus fool other peers to provide an amount of data, before locally being detected as a leeching attempt and switching to a new identity (ID/IP change). Leeching is often observed in situations where peers are unable to upload due to technical limitations. ¹⁶⁰

In order to counter this, a couple of different strategies are adopted. An automated ratings system can be used to artificially boost the priority of verified peers, however, it proves ineffectual against transient leeching, and is best suited for a closed peer group with a low rate of change. This also assumes that peers behave rationally and have a knowledge of the ratings systems in use, such that they can balance their own gain with their overall contribution. This may not always be the case in a torrent network, since common users typically are not aware of the load balancing mechanisms involved. ¹⁶¹

Micro-payment schemes exist as well, which show significantly better results. They are implemented as an extension on top of the torrent framework, but this requires a dedicated third-party handler in the tracker, which puts extra stress on server architecture. This in turn limits the scalability of the system, since each block transaction would go through the central for authorization and accountability. ^{162, 163}

6.4.4 Torrent Availability

-ALEXANDER

As is the nature of a swarm, most “die out” over time, as peers leave the swarm for various reasons. In the case where even the original seeder leaves the swarm, the torrent can become dead. This effectively means that the content of the swarm is now impossible to recover should a new peer connect to the swarm.

The **BT** protocol does not provide a method for revoking torrents with 0% availability and instead relies on the third party sites hosting the torrent metadata files for this. This effectively means that while a torrent site can boast a library of millions of torrents, the actual amount of usable torrent swarms can be significantly lower.

6.5 Peer to Peer

-ALEXANDER

As a content delivery mechanism, any technology based on **BitTorrent** will thus inherit any present security vulnerabilities, unless specifically mitigated against. This in turn means that should the integrity checking mechanisms of **SHA-1** prove insufficient, any additional integrity check need to be introduced and thus requiring a complete overhaul of both checking mechanisms as well as repackaging of existing data and potentially re-verifying already distributed content. This in turn can cripple or at least degrade the performance of business relying on rapid content distribution as a revenue model^{164 165}.

¹⁶⁰ Wang et al., “A Misbehavior Resilient Cipherblock Trading Protocol in BitTorrent-like Networks.”

¹⁶¹ Lai et al., “Incentives for Cooperation in Peer-to-Peer Networks.”

¹⁶² Vishnumurthy, Chandrakumar, and Sirer, “Karma.”

¹⁶³ Yang and Garcia-Molina, “PPay.”

¹⁶⁴ “Akamai: Gamers Aren’t P2P Bandwidth Slaves - TorrentFreak.”

¹⁶⁵ “BitTorrent Goes Legit with Content Delivery Service - InternetNews.”



6.6 End to End

-LARS

June 3rd 2014 Google announced *End-To-End*, (**E2E**) which “implements the OpenPGP standard, **IETF RFC 4880**”.^{166, 167}

End-To-End is a GitHub project with the goal of making it easy to send and receive encrypted e-mail which has been a problem for openPGP. While the size of the core userbase (the so called *strong set*) is rising the overall number of users is rather low, only counting 57'000 users in the *strong set*¹⁶⁸, which dwarfs in comparison to the nearly 1 billion monthly active users of WhatsApps¹⁶⁹.

While it is stated that **E2E** (still under public review) follows **RFC 4880**, it actually does not.

As mentioned in chapter 5.5 openPGP, page 40 **SHA-1** is mandatory for openPGP, but **E2E** does not and defaults to **SHA-256** rather than **SHA-1**^{170, 171} (**LINE 272-273 & 483-514**).

Having launched in 2014, with the code still under public review and with a high level of technical feedback from google employees to the community¹⁷², rather than just following one **RFC** to the letter, security has been improved through public scrutiny and evaluation of newer algorithms.

¹⁶⁶ “Making End-to-End Encryption Easier to Use.”

¹⁶⁷ “Google/end-to-End.”

¹⁶⁸ “Analysis of the Strong Set in the PGP Web of Trust.”

¹⁶⁹ “Facebook’s WhatsApp Hits 900 Million Users, Aims for 1 Billion.”

¹⁷⁰ “Google/end-to-End - Source Code Search for SHA.”

¹⁷¹ “Google/end-to-End Userid.js.”

¹⁷² “S2K Uses Small C/bytcount, Inconsistent Suite of S2K-KDF-SHA1 (160b) and AES-256 · Issue#139 · Google/end-to-End.”



PART 3 PROBABILITY ASSESSMENT

In order to find the Risk related to the Hazards identified in the previous part, it must be combined with a probability.

This part will quantify the probability of finding any type of **SHA-1** collision and in the end give examples of non-**SHA-1** related hazards to compare the threat of **SHA-1** collisions with existing attack vectors.

7 GPU SHA-1 Collision Probability Estimate

-ALEXANDER

To test the hypothesis of whether the **SHA-1** algorithm can be brute-forced, a custom implementation has been made to optimize the process for a specific case: an X.509 certificate.

Fundamentally it has the goal of testing two related aspects of **SHA-1**, general collision resistance and 2nd pre-image resistance, i.e. can a **SHA-1** value be generated to collide with a previously known **SHA-1** value and can two values be generated from different input to produce equal digests.

Since the available computing clusters (see chapter 7.2.2 Available Hardware) feature multiple NVIDIA tesla K40 cards, it makes sense to structure the test program to take advantage of this fact.

7.1 Design considerations

-ALEXANDER

When building the application for the **GPU** a couple of considerations has to be made, which differs from the standard design principles employed on a **CPU**.

Here emphasis is placed on utilizing fast memory types (those closest to the individual thread executors) instead of algorithmic efficiency, since memory access times is a considerable factor in performance. Furthermore, concurrent threads in a specific warp can transfer register contents between them through the built in warp shuffle instructions, but any out of warp memory I/O becomes expensive since it passes through both the level one and two cache to the **GPU** global memory for first a read and then a write operation.

Access to the **GPU** main memory, is cache accelerated and should be done in a consecutive manner if at all possible in order to minimize cache misses.¹⁷³

7.1.1 HPC Forcer Architecture

-ALEXANDER

The main objective of the **HPC** forcer is to generate a lot of SHA-1 values. As seen in Figure 23 Flow of **HPC** forcer implementation and Figure 24 Flow of **HPC** forcer implementation **GPU** (Device) side, there are several steps to this process.

The host program has to acquire handles to the **GPGPU** devices and then generate the input to the kernels before ultimately launching the kernels, before capturing their results and returning it to the user.

The device code computes a hash value and then compares it in order to determine whether a meaningful result was created.

¹⁷³ "CUDA C Programming Guide."



In both the case of the client side and host side code, some or more of these steps are executed several times in order to maximize the time spent generating and comparing **SHA-1** instead of transporting data and setting up devices.

All of this is implemented in the Sha1.cu file (source code can be found in Appendix).

Actually deploying the executable is done by a script which interfaces with the HPC queue deployment system.

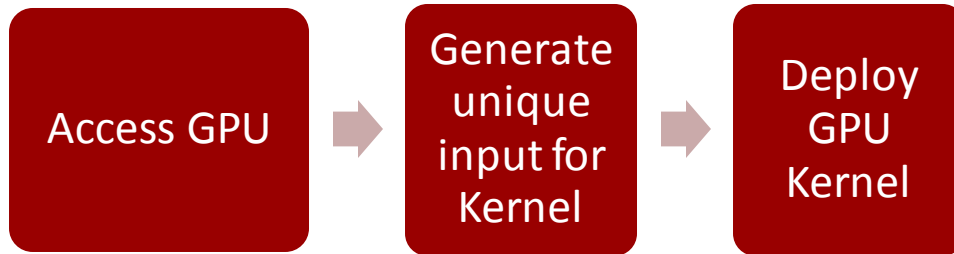


Figure 23 Flow of HPC forcer implementation CPU(Host) side

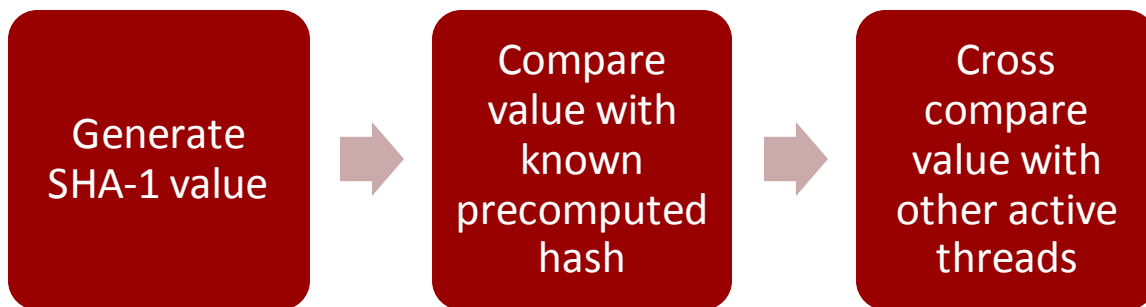


Figure 24 Flow of HPC forcer implementation GPU (Device) side

7.1.2 SHA-1 Kernel

-ALEXANDER

The kernel is defined in the global function “SHA1Kernel” and takes as input a random seed and a counter variable. The random seed is generated by the host when starting up and it helps differentiate the hosts and ensures that no two sets of kernels launched by the host are similar.

The counter variable allows the kernel to start from a new section for each invocation and helps differentiate the input between invocations.

The kernel starts by setting up the principal **SHA-1** state variables h0 through h4. They are however set to the initial value precomputed for the target certificate instead of the default SHA-1 initialization vector (see the Sha1.cpp file in appendix for the IV precomputation).

This is different from the normal SHA-1 implementation where the initial state is set to: H0->h4 = 0x67452301, 0xEFCDAB89, 0x98BADCFE, 0x10325476, 0xC3D2E1F0.

The input block of 512 bits is kept in the chunk array in the first 16 slots. This input is generated in the “initializeChunk” function, which takes the input variables together with a thread id and a kernel counter to generate a unique chunk for each iteration.

The main “SHA1Chunk” function is then called to calculate the digest, which is stored in the state variables h0 through h4. This is implemented in two segments. The first part consists of the input expansion which is the loop that generates the remaining parts of the input block slots 16 through 79. The second part consists of 80

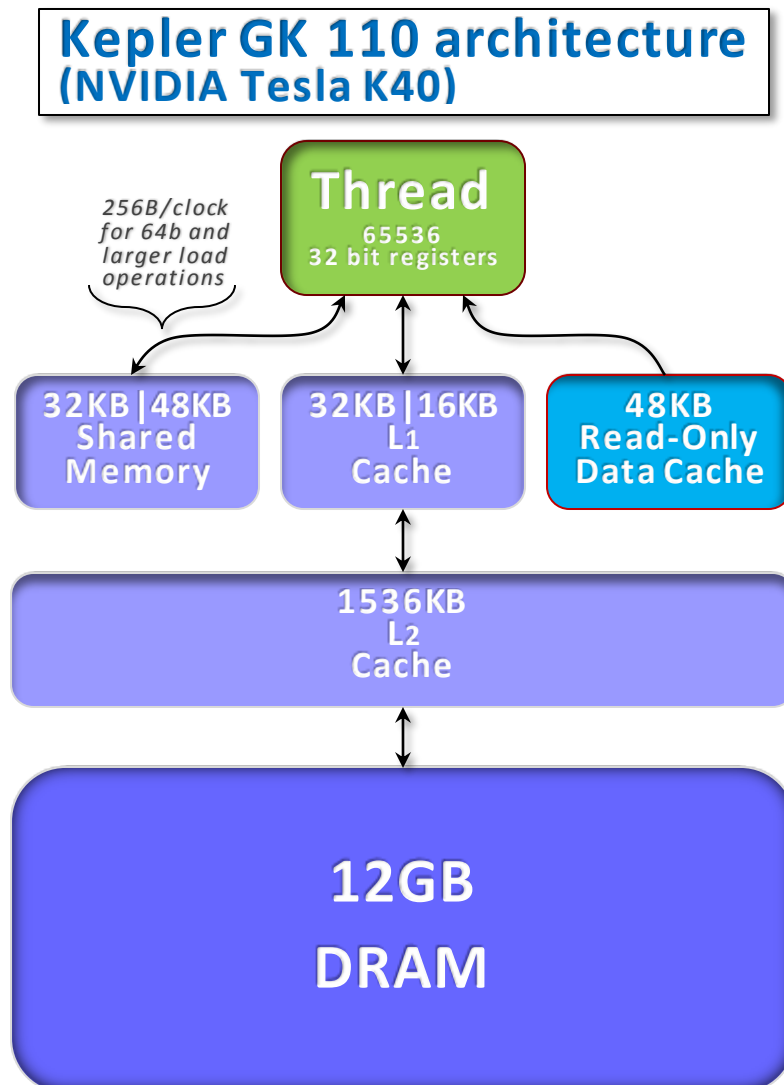


SHA-1 rotations, which internally modify the temporary state variables a, b, c, d and e according to the **SHA-1** specification.

Finally, the computed variable is compared to the target vector and to each other vector in the thread block by means of the new shuffle instructions. And in the case of any type of collision the input chunk and the output values is returned to the host thread.

This output is however the absolutely only time the kernel returns any value. As it is an expensive operation and too much output will break the amount of memory space available to a **CUDA** kernel. But in highly unlikely case of any **SHA-1** collision, all performance considerations are irrelevant and the output is returned directly.

Of special note is the fact that every function beyond the global entry point is kept inline, and a very sparing amount of temporary memory is used. This is done in an attempt to keep memory used within registers and the level 1 cache, in order to not impact performance due to memory delays. And while the **GPU** will swap between warps during waiting periods, this operation still takes time and as such is best avoided if possible.



Based on¹⁷⁴, by Lars Embøll

¹⁷⁴ NVIDIA, "Kepler Compute Architecture Whitepaper."



This kernel design thus allows for testing both general and 2nd pre-image collisions at once and assuming that the cross comparison time is negligible in performance impact, it will do so in at the rate of which **SHA-1** digests can be generated.

7.1.3 Optimizations

-ALEXANDER

To improve performance, the chunk expansion loop is removed and replaced with an unrolled version, where a lot of the steps have been removed. This is allowed because we know the exact form the input takes. And can therefore remove a lot of the steps which can be precomputed. This removes the equivalent of 3-5 steps of the loop.

A second improvement is in the outer 80 rotations which have been split up into 4 loops of twenty steps each. This removes 4 control statements for each loop which in turn obviates the need for any branch prediction during the inner loop. The four inner loops are then unrolled by the compiler to remove the flow control statements inherent in the for-construction.

The result is that the **SHA-1** calculation feature no control statements which should hopefully speed up execution.

7.1.4 Prehash Value

-ALEXANDER

In order to speed up the 2nd pre-image collision generation process, only the absolute minimally necessary amount of input is hashed for each collision-generation attempt. Since **SHA-1** operates on 512bit input blocks, any input larger than that will have to be split up in two or more sections of 512 bits and each individual block will be fed to the **SHA-1** inner algorithm in order.

Between blocks, the intermediate state is stored in the initialization vector variables h0 through h4, and by making sure only the very last input block changes between collision attempts, the preceding digest values can be computed once and stored for each subsequent collision attempt.

This updated initialization vector is called the prehash value and by utilizing that as the new initialization vector, 2nd pre-image collision attempts will take only one **SHA-1** attempt, no matter the size of the input, as long as the total size of the terminating blocks is no more than 440 bits (512 bits minus the final **SHA-1** padding) and it is aligned with the 512bit boundary.

7.2 GPGPU & CUDA

-ALEXANDER

While the classic **CPU** based programming model is effective at solving many general programming tasks, utilizing specialized hardware such as graphics cards is often preferable when the problems presented can be parallelized on a massive scale. This type of problems includes areas such as finite element simulations, linear algebra and image rendering.

The task of computing a single hash value is by nature impossible to parallelize due to the avalanche effect¹⁷⁵, and as such it is ill suited for **GPU** work. However, the task of computing multiple different values is perfectly suited for **GPU** work, due to the fact that there is no intra dependencies between computation threads, beyond making sure that the input is distinct for each attempt.

¹⁷⁵ Stafford E., *On the Design of S-Boxes - Advances in Cryptology*.



When developing code for **GPU**s, there are several languages and framework choices, but since there are NVIDIA **GPU**s available, the **CUDA** framework/language is used as it natively executes on any modern NVIDIA card.

7.2.1 Language Variations

-ALEXANDER

CUDA implementations comes in several language variants and abstraction layers¹⁷⁶, each with their own focus and target language, including python, C/C++, LUA and more. However only the bare toolkits are guaranteed to be supported on any **CUDA** setup which makes the raw C/C++ versions the most reliable development choice in terms of portability assurance.

The bare framework is the **CUDA C/C++ SDK**, built by NVIDIA with the NVCC Compiler based on LLVM¹⁷⁷. It offers full C++ functionality for host code and most of the C++ functionality for device/kernel code as well. Furthermore, reference code exists, which implements a similar type of computation in **CUDA C/C++**¹⁷⁸, further increasing the likelihood of a successful implementation.

7.2.2 Available Hardware

-ALEXANDER

In cooperation with **DeIC**¹⁷⁹, a number of computing nodes have been made available, featuring the NVIDIA Tesla K40 cards, with support for the **CUDA** 3.5 environment, provided by the GK110 Kepler chip architecture¹⁸⁰. These are available in the ABACUS 2.0 **GPU** cluster specifically designed and optimized for high-throughput number crunching, however the computational power of the GK110 is optimized to be able to execute many concurrent threads at the same time, with some restrictions on the type of job they can handle. Each individual thread executes at a variable, but rather low frequency, typically in the area of 700 Mhz and in addition to this the execution units lack some modern features typically found in **CPU**'s such as branch prediction. However, instead they make up for this fact by having several concurrent multiprocessors with the ability of rapidly switching between multiple threads in order to eliminate stalling and thereby effectively being able to execute every instruction in one clock cycle.

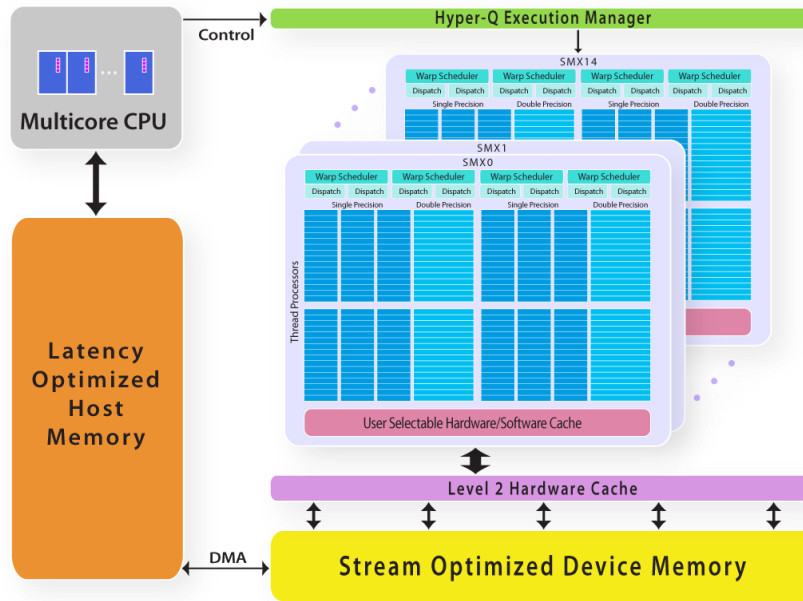
¹⁷⁶ "Language Solutions."

¹⁷⁷ "NVCC :: CUDA Toolkit Documentation."

¹⁷⁸ "Cryptohaze.com • View Topic - CUDA Multiforcer 0.7 Source."

¹⁷⁹ "Abacus 2.0 | DeIC National HPC Centre, SDU."

¹⁸⁰ NVIDIA, "Kepler Compute Architecture Whitepaper."



©2013 The Portland Group, Inc.

Figure 25 Parallelization architecture of the GK110 chip. The introduction of the Hyper-Q Execution Manager allows for multiple applications to utilize the chip in greater effect¹⁸¹.

7.2.3 Core concepts

-ALEXANDER

This section will highlight the core concepts needed to understand, develop and deploy a GPU application.

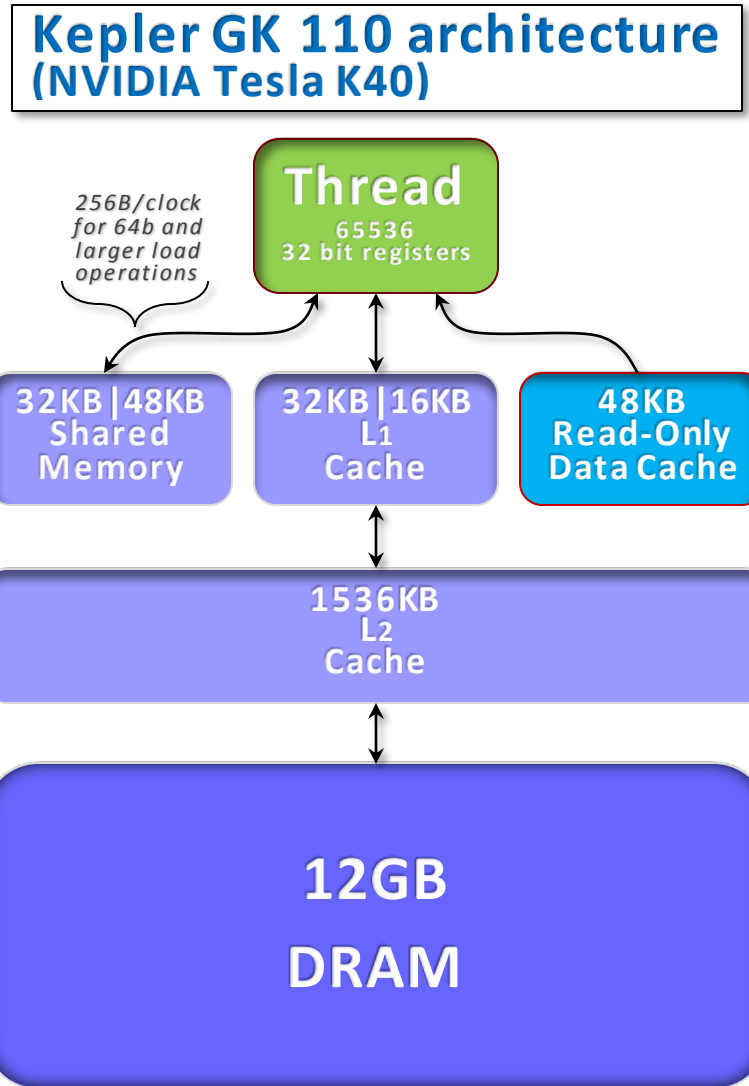
Streaming Multiprocessor (SMX):

This is analogous to the CPU; the unit that will execute kernel code. It contains several layers of memory from individual thread registers to shared texture memory and level 1 cache. It stores many grid blocks of threads internally and switches between executing these in order to maximize throughput. The GK110 features 15 SMX units where each SMX is equipped with 192 execution cores, as seen in the specs in Figure 28 SMX block diagram.

	FERMI GF100	FERMI GF104	KEPLER GK104	KEPLER GK110	KEPLER GK210
Compute Capability	2.0	2.1	3.0	3.5	3.7
Threads / Warp	32				
Max Threads / Thread Block	1024				
Max Warps / Multiprocessor	48		64		
Max Threads / Multiprocessor	1536		2048		
Max Thread Blocks / Multiprocessor	8		16		
32-bit Registers / Multiprocessor	32768		65536		131072
Max Registers / Thread Block	32768		65536		65536
Max Registers / Thread	63			255	
Max Shared Memory / Multiprocessor	48K				112K
Max Shared Memory / Thread Block	48K				
Max X Grid Dimension	2 ¹⁶ -1		2 ³² -1		
Hyper-Q	No			Yes	
Dynamic Parallelism	No			Yes	

Figure 26 Compute Capability of Fermi and Kepler GPUs

¹⁸¹ Pgi-Kepler-Block-Diagram.png (PNG Image, 1152 x 864 Pixels) - Scaled (79%).



Memory allocation in the Kepler GK 110 architecture, by Lars Embøll, derivative of ¹⁸²

Threads: They are parallel to, and roughly equivalent to a **CPU** thread. They each run a single **GPU** function (called a Kernel) at a time. Depending on the implementation-language and implementation environment, they can do most of what a normal thread is able to do. They are however typically very limited in individual (local) memory available, and therefore rely mostly on shared memory for large datasets.

Kernel:

A **GPU** program. Much like an arbitrary **CPU** program, but with some restrictions. While **CPU** programs can return a value, **GPU** kernels cannot, since many thousands can run at the same time and it would not be clear where to return the value to.

Kernels cannot throw exceptions, since the overhead associated with exceptions would be catastrophic in a massively parallel context.

Each Kernel has a limited amount of registers, for local storage, which in general means that kernels should not rely on local data except for control flow. Note that constants (embedded in the program code) and input/output vectors do not count in this limitation as they are visible across the entire **GPU** space.

Thread Block:

Groups of threads are deployed to a **GPU** as thread blocks. They have a shared memory space for internal communication which typically resides in shared memory. Each thread block is wholly deployed to one **SMX**

¹⁸² NVIDIA, "Kepler Compute Architecture Whitepaper."



unit, which means that in order to fully utilize a **CUDA** device, at least as many thread blocks as **SMX** units must be deployed.

Grid: A grid is a collection of thread blocks, and is what is launched together with a given Kernel in order to process a given problem. For a simple **GPU** program, it is typically enough to launch one of these and the **GPU/CUDA API** will distribute the thread blocks to the individual **SMX** units.

Warp:

Within a **CUDA** device, the processor executing instructions does so across a number of threads at a time. It performs the same instruction across a warp, which is a grouping of individual thread lanes.

While each thread features its own registers, it does not feature its own program counter, which is instead synchronized across the entire warp. When a warp stalls (typically for I/O bound actions), another warp will execute in its stead. **CUDA** has a set limit of 32 threads per warp and within a warp, there exists special instructions for high speed intra-thread communication.

Occupancy:

The theoretical maximum active threads in a chip is defined as the amount of **SMX** units times the threads/warp, however each **SMX** has a limit on available memory, so the amount of threads active can need to be scaled down in order to fit the memory footprint (the ratio between max and utilized threads is called the occupancy and the higher the occupancy, the more raw data is being processed).

The Occupancy metric is important since the higher the occupancy, the more options are available to the **GPU** for effectively managing its workload.

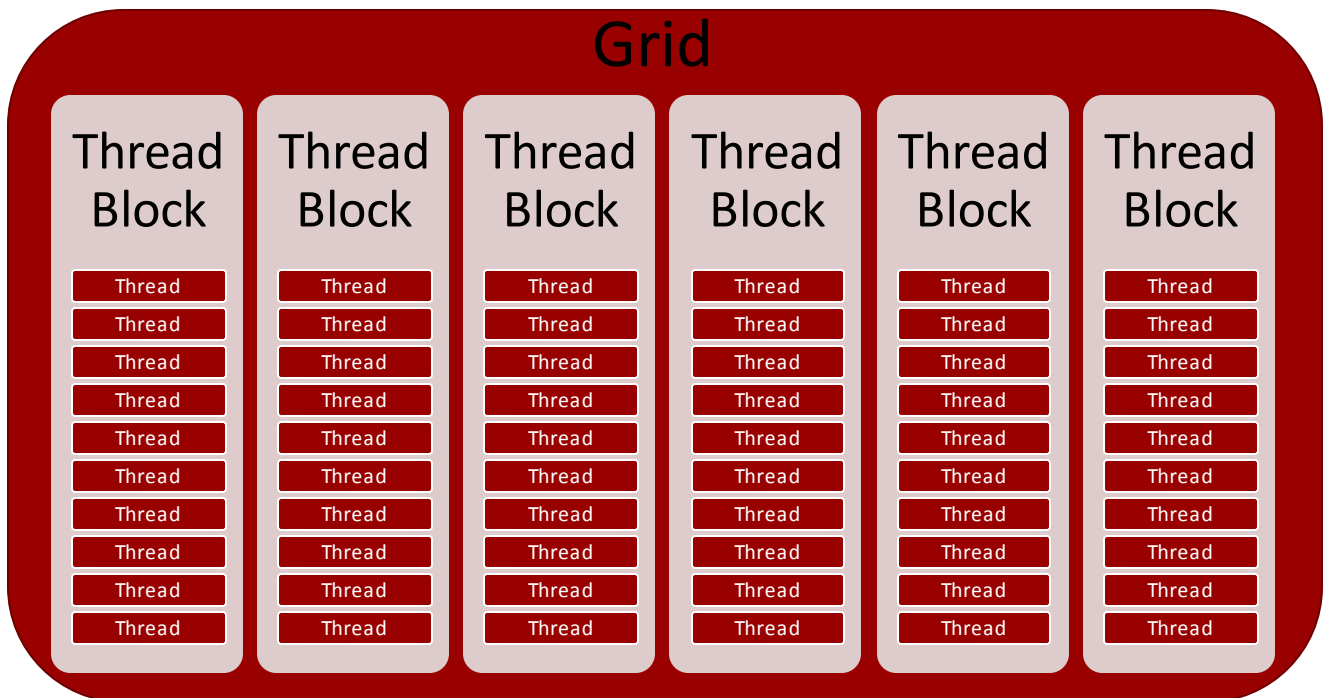
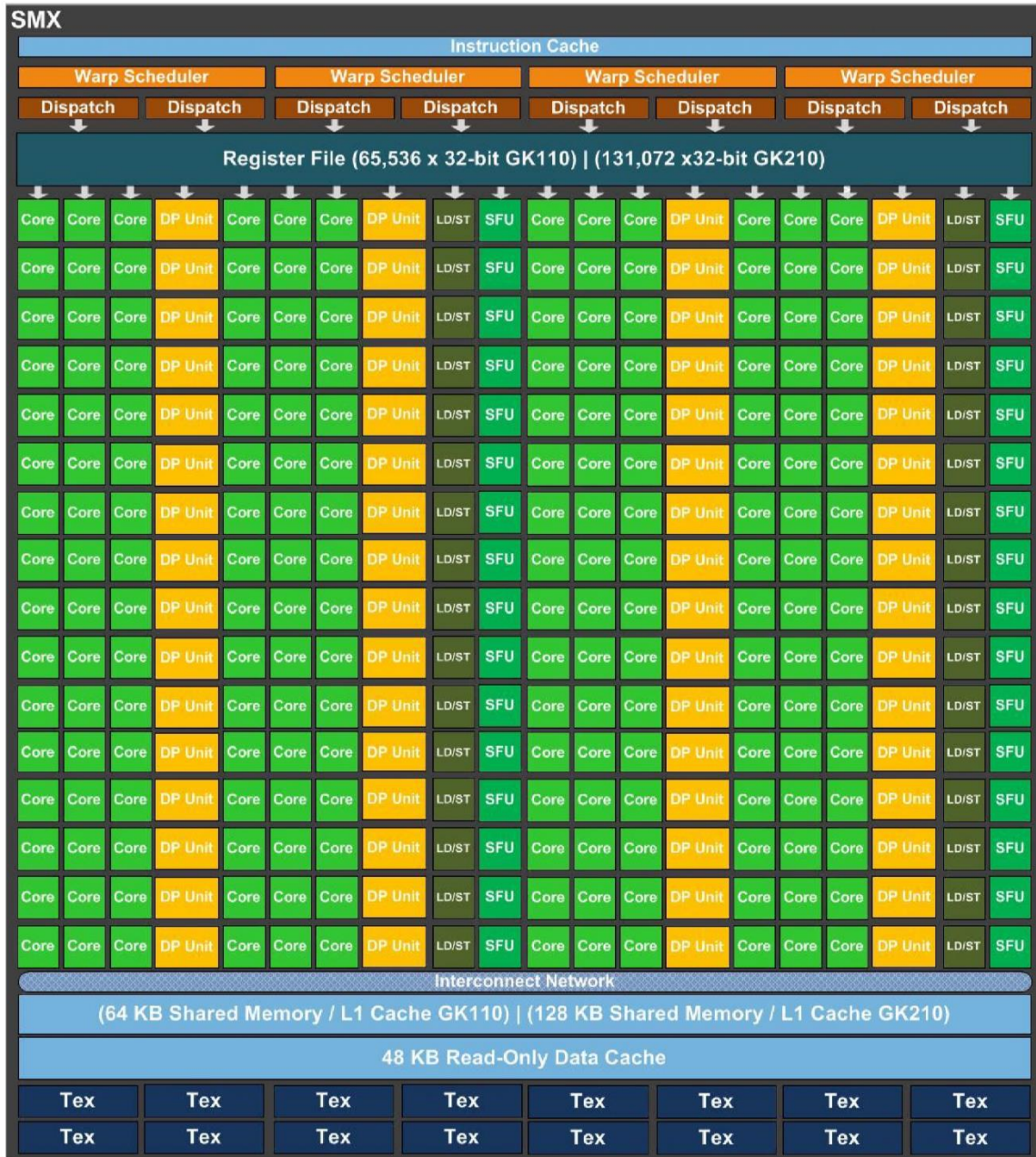


Figure 27 Diagram illustrating the relationship between threads, threadblocks and grids



SMX: 192 single-precision CUDA cores, 64 double-precision units, 32 special function units (SFU), and 32 load/store units (LD/ST).

Figure 28 SMX block diagram

7.2.4 Memory model

-ALEXANDER

The GK110 has memory in 4 layers. Each individual thread features 255 program accessible registers for local storage which is kept in the register file. Beyond that it is connected to a level 1 cache and a chip wide shared memory as well as texture storage area (specialized in rapid read but slow write times, which is typically what is used in 3D texturing). Beyond this, each card has 12GB of global memory which is accessed through a level 2 cache first.

The important thing to note here is that everything beyond level 1 cache access accesses parts of the K40 beyond the SMX and as such is considered extremely time consuming and is therefore best avoided if possible.



7.2.5 CUDA C/C++ specifics

-ALEXANDER

CUDA C/C++ is a superset of C/C++, with added instructions specific for operations between a **GPGPU** and a **CPU**, as well as language constructs for designating methods to be transferred and executed on the **GPGPU** instead of a **CPU**. In the end, a **CUDA** program compiles to a normal executable, with embedded **GPU** code that is deployed to the **GPU** during launch. There exist solutions for **Just In Time** compilation of **GPU** code as well, mainly aimed at the possibility of utilizing better future compilation techniques and better hardware.

Kernel definition:

Defining a Kernel in **CUDA** C/C++ is done just by writing a normal function, but prefixing it with the “`__global__`” keyword. This makes the function visible both the **GPU** and **CPU**, note that the above mentioned limitations still apply to the Kernel definition, and as such not strictly all valid C/C++ instructions will be allowed.

The kernel function can take arbitrary input as arguments, but it is important to remember that any arguments will be copied through the **PCI** bus to each thread and each thread will reserve a copy of the data in either its registers or shared memory. Therefore, it is important to restrict the amount of data a kernel accepts at launch. While “`__global__`” functions are visible from both the **CPU** and **GPU**, kernels can also define their own auxiliary methods. This is done by prefixing them with the keyword “`__device__`”, which implies they are only reachable through other device functions.

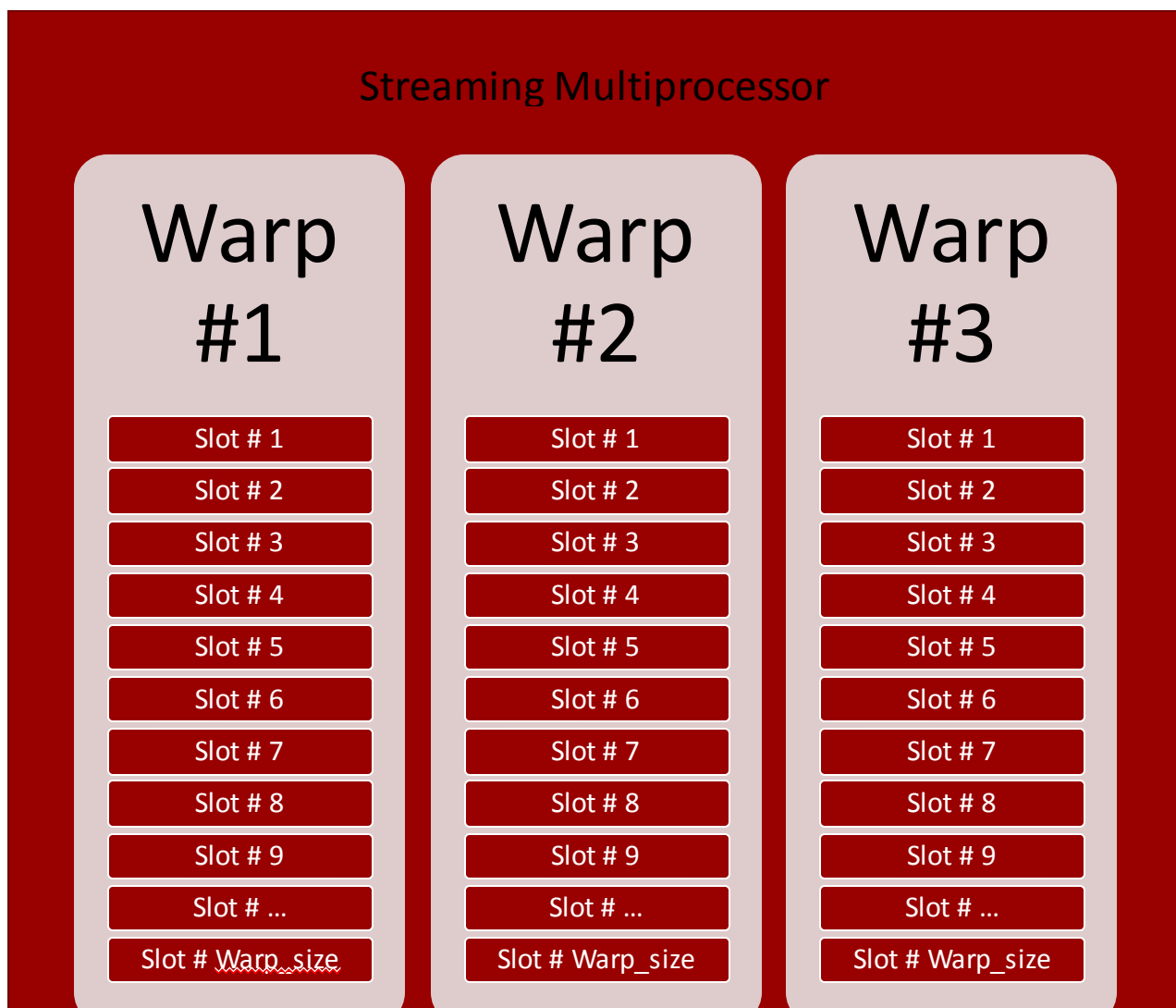


Figure 29 Illustrating the relationship between warps and and the **SMX**.
Warp size is currently defined to be 32 for all known **CUDA** implementations.

**Kernel invocation:**

Kernels are invoked with the triple-chevron syntax “<<<n_blocks,n_threads>>>” syntax, which specify what is called the “launch configuration”, which is essentially the amount of **GPU** threads used simultaneously.

Depending on the available chipset, the maximum n_blocks/n_threads vary, but combined numbers of more than millions total threads in a launch configuration is not impossible, but it is very unlikely that all thread will execute simultaneously¹⁸³. This chevron syntax is a special compiler flag which is translated into **CUDA API** calls at the compilation pre-processing step.

A thread block is run simultaneously, and depending on the specific card in use, multiple blocks can be launched simultaneously. This is device dependent and as such when launching a Kernel, it should be done with the knowledge that it is potentially unknown how many blocks are running at once due to scheduling.

Once a kernel is launched, the **CPU** will resume execution and will not further interact with the kernel until calling the “cudaDeviceSynchronize()” method, which will block until the kernel has finished executing and has returned a status indicator.

Data transfer & Message passing

Transferring data back and forth between the host and device can be done by passing arguments to the launch function. However, for larger amounts of data, that does not need to be duplicated between threads, another method exists.

It operates in much the same way as using the traditional C malloc, but instead provides the “cudaMalloc” function, which allocates an area of **GPU** memory available and returns a **GPU** pointer to this area.

It works in conjunction with the “cudaMemcpy” function, which is capable of copying data to or from a host and **GPU** as well as from **GPU** to **GPU**. It uses the **GPU** pointers provided by the cudaMalloc functions instead of native pointers.

Beyond this mechanism, the **CUDA API** provides a kernel implementation of the “printf” function and its derivatives. A kernel can print a message which will be caught by the host thread upon reaching the synchronization method.

Care should be taken not to overuse this function, as the messages are store in **GPU** memory before transfer and all messages are sent through the **PCI** bus, meaning that a lot of output can destroy kernel performance.

Shuffle instructions

New in the **CUDA 3.X** and newer architectures is the shuffle instructions. They allow the **GPU** to transfer 4 bytes of content between threads in a warp in a single instruction. This is a new addition and allows for the fastest possible intra-thread communication, since no messages pass between busses or is copied to cache/memory. The caveat is that the shuffle instructions only work within a warp and take no consideration to whether the warp is fully utilized (can copy garbage from inactive cores) or allow access to neighbouring threadblocks. These instructions are especially useful for intra thread comparisons and summations, as well as vector/matrix math.

¹⁸³ “CUDA Occupancy Calculator Helps Pick Optimal Thread Block Size - NVIDIA Developer Forums.”



8 BitTorrent

-ALEXANDER

In order to determine the feasibility of utilizing the BitTorrent network to harvest SHA-1 values. A set of test suites are setup in order to extract the maximal amount of SHA-1 data from the torrent network possible.

8.1 Data source

-ALEXANDER

When gathering raw torrent data, there are two options. Either utilizing magnet links or regular .torrent files. In the case of torrent files, there is a need to contact a dedicated indexing site and download all the available torrents, either in the form of scraping the site or by way of a specialized API.

In the case of magnet links, the process is much the same except the need for an extra step in resolving the individual magnet links and converting them into usable torrent files. However, a positive note on magnet links is that they feature a built in availability check, since if a magnet link is unresolvable, there is a high chance of the content torrent being unusable/unavailable as well.

While this comparison makes regular .torrent files somewhat preferable, the most important constraint, is where it is possible to get a large sample of torrents, in order to gather proper statistics.

Here it is favorable to have access to more than a million individual torrent records and in the end, the most significant factor, when choosing the type of input to parse is simply availability.

Here any solution which relies on scraping of indexing sites must be considered a last resort as it is extremely error prone as well as only guaranteeing a snapshot of a specific indexing site instead of a more comprehensive view of the torrent ecosystem.

Luckily, the BitSnoop site ¹⁸⁴, offers a daily scan of the majority of the torrent ecosystem in a compressed format. And this will form the basis of the input data used.

Through its published API it provides access to a complete dump of its indexed torrent, which are collected from a wide array of sources. These sources are claimed to be at least a thousand different trackers, and while it is highly unlikely for torrents to only exist on one tracker, Bitsnoop does a rudimentary check to verify the integrity of individual torrents and then filter out possible duplicates.

While this process is far from foolproof, it provides a good starting point as a data source. Which makes it highly desirable.

8.2 BitSnoop Data extraction

-ALEXANDER

The torrent data offered, is in the form of a giant list of all its indexed torrents, aggregated in one file.

Each record has its own line (unix line breaks) with pipe separated fields. The two first fields in a record are the most important ones, being info hash and name, which is all that is needed to extract a functioning magnet link. The remaining fields are categories, download URL and info URL which are used for keeping track of where the torrents originate from.

This data is parsed and converted into a list of magnet links, which is then sorted by info-hash and any duplicates are discarded since a magnet link resolver would not be able to tell a difference.

¹⁸⁴ "About Us | Bitsnoop."



8.3 Magnet link resolution

-ALEXANDER

In order to gather usable statistics on individual torrents, the magnet links need to be converted into torrent data descriptors. Normally this is done by following the described magnet link protocol extension, which would terminate when a magnet link is either resolved or it is determined that the specified torrent is no longer contained in any active swarm (it is a dead torrent).

An alternative is to utilize a third party caching site and resolve the magnet links through that, such as the one offered through Torcache¹⁸⁵.

Ideally, the two solutions should be combined, and there is no reason to exclude one solution over the other.

For raw throughput, the dedicated cache will be expected to outperform the manual magnet resolution protocol, since it potentially involves multiple connections, where each could target any place on the globe.

For this reason, the resolver program will first attempt to resolve links by the cache mechanism and in the case there are unexpected performance penalties, the fallback mechanism will be manual magnet link traversal.

9 When will we see a SHA-1 collision?

-LARS

An estimate saying that **SHA-1** (general) collisions would cost **USD** 700K in 2015 was made by Bruce Schneier October 2012¹⁸⁶.

This estimate relies on a few key factors:

- A. 2^{14} cycles per **SHA-1** block
- B. 2^{60} reduced space for **SHA-1**¹⁸⁷
- C. Doubling of **CPU** power each 2nd year. (Moore's law)

With Schneier's predictions being part of the foundation for the decisions by Microsoft¹⁸⁸ and Google¹⁸⁹ to deprecate **SHA-1**, the following section will try to give an updated evaluation of those assumptions.

A) Cycles per **SHA-1** block (64 byte/512 bit)

By using **GPU** rather than **CPU** a much more optimized instruction set can be used, resulting in a decreased number of cycles needed for each block.

Even with **CPU**, such as 2015 Intel Core i5-6600; 4 x 3310MHz it is reduced to 4,32 cycles/byte¹⁹⁰, for sizes > 576 bytes, down to 3,60 for sizes > 4096 bytes. (and up to 10,81 for messages of 64 bytes)

Compared with Schneier's $2^{14} / 64$ byte block = 256 cycles / byte.

Improvement: Factor 59-71 (23)

B) Outcome space of **SHA-1**

Massively parallel architectures¹⁹¹ (page 17-21) give the possibility to exploit the birthday weakness.

Improvement: 3-4¹⁹², only viable for general and not 2nd pre-image collisions.

¹⁸⁵ "Torcache - Torrent Cache."

¹⁸⁶ Schneier, "When Will We See Collisions for SHA-1? - Schneier on Security."

¹⁸⁷ Stevens, "Cryptanalysis of MD5 & SHA-1."

¹⁸⁸ "SHA1 Deprecation Policy - Windows PKI Blog - Site Home - TechNet Blogs."

¹⁸⁹ "Intent to Deprecate: SHA-1 Certificates - Google Groups."

¹⁹⁰ "Measurements of Hash Functions, Indexed by Machine."

¹⁹¹ Stevens, "Cryptanalysis of MD5 & SHA-1."

¹⁹² NVIDIA, "Kepler Compute Architecture Whitepaper."



C) Moore's law

As seen in chapter 3.4, page 31 Moore's law is stagnating, while having had the general trend of doubling each second year for over 4 decades, it recently has only been doubled each two and a half years¹⁹³.

Improvement: -¼ less per year

The following chapter will give an update to these values in order to give a better estimate for the arrival and price of the world's first **SHA-1** collision.

10 SHA-1 Collision Testing

-ALEXANDER

This section is dedicated to the results of the **SHA-1** tests, trying to quantify the degree of which any **SHA-1** collisions can be found, as well as characterize the amount of resources needed to be invested in order to get one.

10.1 HPC Diagnostics

-ALEXANDER

The diagnostics raw data can be found in the files: "diagnostics-133144.out", "GPU0.txt" and "GPULIST.txt" that have been submitted with this report and they are the result of running the diagnostics slurmscript.

Beyond confirming that each node operates with two standard Tesla 40K cards, their power consumption limit under a full load (which our test program draws), is 235 Watt per card.

The cards are kept in a default configuration, with no overclocking and no automatic clock upscaling (the clock will not increase in frequency during peak loads).

No resource sharing is present so any process submitted can be assumed to have full access to the entire card.

10.2 HPC SHA-1 generation

-ALEXANDER

Due to the **HPC** cluster being a shared resource, it was not possible to run on all 72 **GPU** nodes at the same time without a delay of at least 2 days, if dynamic load was selected, the jobs got processed immediately.

The resulting **SHA-1** performance is as follows; it is important to note that neither version managed to generate **SHA-1** collisions of any kind.

Normal SHA-1 Kernel	
devices	2
Kernels	32
Blocks	32'768
Runs	1'000
Lines	554
SHA-1 Values	5,80911E+11
Time (Seconds)	4'320

¹⁹³ Clark, "Intel Rechisels the Tablet on Moore's Law."



SHA-1/Second	134'470'163
SHA-1/Second/card	67'235'081
Watt/card	235
SHA-1/Joule (Watt/s)	286'000

Optimized SHA1 Kernel	
Devices	2
Kernels	32
Blocks	32'768
Runs	1'000
Lines	1'380
SHA-1 Values	1,44703E+12
Time (Seconds)	4'320
SHA-1/Second	334'961'778
SHA-1/Second/card	167'480'889
Watt/card	235
SHA-1/Joule (Watt/s)	712'000

10.3 HPC Evaluation

-ALEXANDER

The peak performance of the HPC setup was 167'480'888 **SHA-1** values per second per card for the optimized version of the code.

Extrapolated across all **GPU** nodes this equates to a maximum throughput of 24'117'247'987 **SHA-1** values per second, which in raw numbers is factor 7 increase in throughput compared to the amazon cloud implementation from 2011¹⁹⁴.

While indicative of raw power, this comparison does not offer a configuration independent performance metric, therefore a much more interesting metric is the amount of hashes per Watt and subsequently the price per produced digest, based on the power consumption.

The performance of a single card is 713'000 SHA-1/Joule, which means an investment of one MWh will return 2'566.8 Tera-hashes.

At the time of writing, the fixed power price, offered to Danish home users by DONGENERGY is: 0,2633 **DKK/kWh**¹⁹⁵, which can be used to derive the price of generating collisions.

This will be contrasted with the price of doing the same calculation by utilizing the Amazon Elastic cloud. Here, the GPU nodes fitted with GK104¹⁹⁶ chips are used, which are sufficiently close in architecture to allow a comparison. They are available on the G2.2xlarge **GPU** instance at a price of 0.65 **USD** per node hour¹⁹⁷.

¹⁹⁴ Roth, "Cracking Passwords In The Cloud."

¹⁹⁵ "Elpriser – Se de Aktuelle Elpriser Hos DONG Energy."

¹⁹⁶ "Product Details."

¹⁹⁷ "EC2 Instance Pricing – Amazon Web Services (AWS)."



HPC Application results		
For each MWh invested, Chance of success		
Second pre-image attack (X.509 Forgery)	1.756×10^{-33}	
Any kind of SHA-1 collision	2.809×10^{-32}	
Digests needed to reach 50% Probability		
Second pre-image attack (X.509 Forgery)	1.008×10^{48}	
Any kind of SHA-1 collision	6.302×10^{46}	
Price of successful collision by power (Exchange rates based on 1/1-2015)		
Second pre-image attack (X.509 Forgery)	1.034×10^{35} DKK	1.488×10^{34} USD
Any kind of SHA-1 collision	6.468×10^{33} DKK	9.313×10^{33} USD
Price of successful collision by Amazon Cloud		
Second pre-image attack (X.509 Forgery)	1.087×10^{36} USD	
Any kind of SHA-1 collision	6.794×10^{34} USD	

10.4 Applied Pigeonhole

-ALEXANDER

Assuming the **GPU** application was capable of storing the digests generated without it affecting its performance significantly, the probability of producing general collisions would be dramatically altered. The amount of hashes needed to be generated in order to make the chance of a general collision above 50% would be 2^{80} digests. Factoring in current weaknesses found by Marc Stevens¹⁹⁸, this is pushed down to 2^{60} .

This is considerably closer to a realistic goal, but it does however completely ignore the storage aspect, which would likely have a tremendous impact on the generation rate.

And while this can maybe be realistically accomplished, the search time through a database of that size would dwarf the time used to generate the hash values in the first place if it were to be performed by the **GPU** thread.

Extrapolated HPC results (Including Pigeonhole effect)		
Price of successful general collision by power (Exchange rates based on 1/1-2015)		
General SHA-1 collision (full)	1.240×10^{11} DKK	1.785×10^{10} USD
General SHA-1 collision (reduced)	1.183×10^5 DKK	1.703×10^4 USD
Storage requirements		
General SHA-1 collision (full)	4.352×10^{13} Terabytes	
General SHA-1 collision (reduced)	4.150×10^7 Terabytes	

¹⁹⁸ Marc, "Cryptanalysis of MD5 & SHA-1."



10.5 BitTorrent SHA-1 Extraction

-ALEXANDER

Approximately 23 Million torrent descriptions were offered by the **API**, and all of them were fetched in one large file.

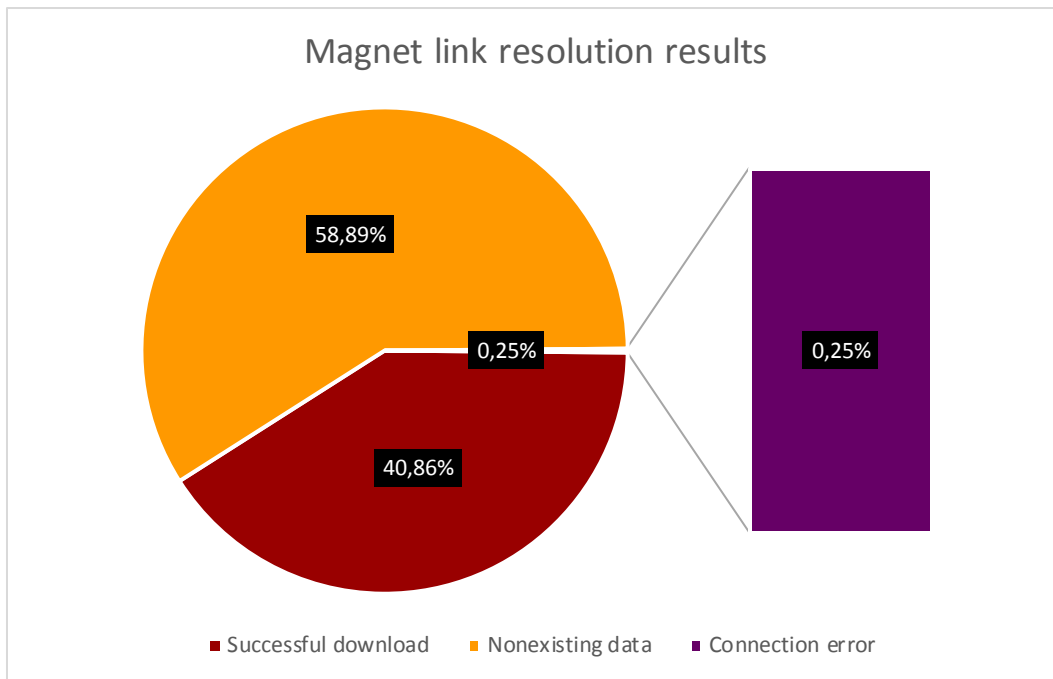
This was parsed into unique magnet links, sorted and then any duplicates were removed based on the info hash. From there the links were resolved and the end result was 5'926'664 unique magnet links were extracted and downloaded over a two-week period, describing unique torrent files that currently or have previously existed. This process was stopped prematurely and by linear interpolation, it would have taken another estimated 2 weeks to complete fully.

Even with the quality control offered by Bitsnoop, this sample featured only 25% unique descriptors with the rest being duplicates in some form.

The remaining magnet links were then resolved against the torrent database with the following result:

Result distribution	Count
Total Exceptions:	3'505'280
Total Torrents:	2'421'384
Error source	Count
The remote server returned an error: (404) Not Found.	3'490'257
The CRC in GZip footer does not match the CRC calculated from the decompressed data.	4
The remote server returned an error: (502) Bad Gateway.	91
The operation has timed out.	14'646
Unable to connect to the remote server ---> An invalid argument was supplied 95.215.61.199:80	23
Unable to connect to the remote server ---> A connection attempt failed because the connected party did not properly respond after a period of time, or established connection failed because connected host has failed to respond 95.215.61.199:80	10
The remote server returned an error: (500) Internal Server Error.	246
Unable to connect to the remote server ---> A connection attempt failed because the connected party did not properly respond after a period of time, or established connection failed because connected host has failed to respond 109.163.226.148:80	2
The request was aborted: The operation has timed out.	1

By aggregating the error sources, the following graph illustrates the result.



Where any error beyond the non-existing data error, is insignificant in comparison with the size of the data.

The downloaded .torrent files feature these characteristics in terms of contained **SHA-1** blocks and their respective block sizes:

There were a total of 3'287'001'641 chunks present, which in other words is the number of distinct **SHA-1** digests.

They originated from a total of 2'407'214 torrents counted, with an average amount of chunks in each file being 1'365.

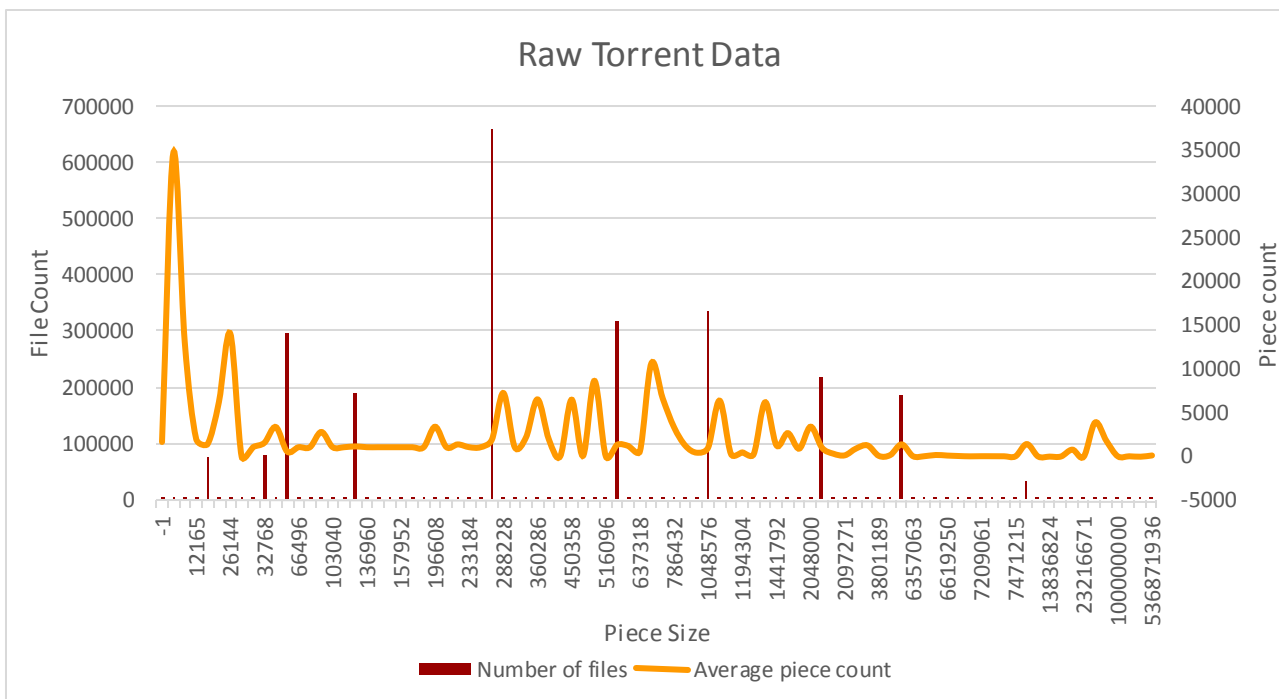


Figure 30 The total raw Torrent data. The x axis denotes the piece size, the left y axis the file count and the right axis the piece count.

By purging the torrents containing an irregular piece size and specifically keeping the 10 most common counts the same graphs can be obtained, but with more usable data.

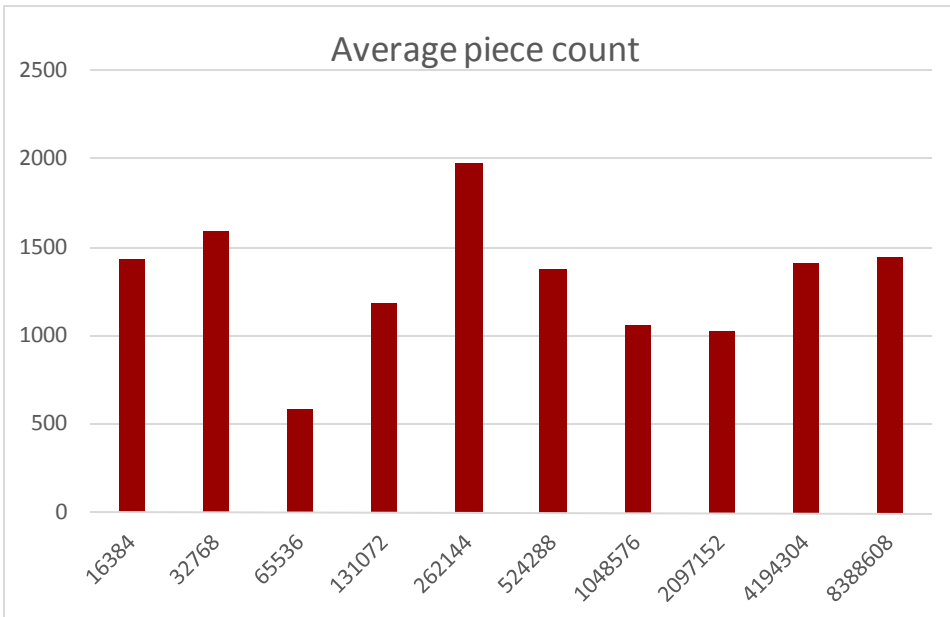


Figure 31 Piece count in torrent files. The x axis denotes the piece size and the y axis the piece count, outliers removed.

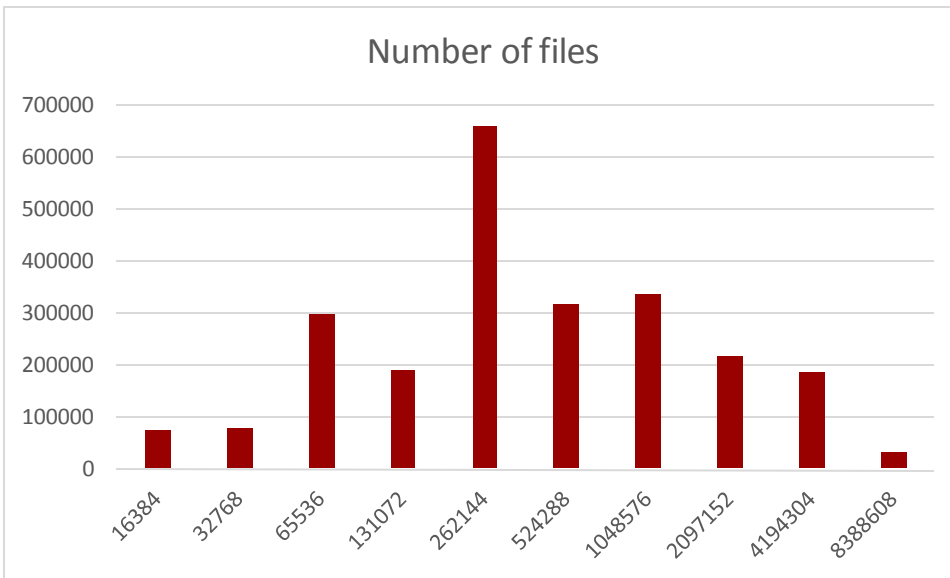


Figure 32 Amount of torrent files with the specific piece size. The x axis denotes the piece size and the y axis the file count, outliers removed.

As a side note, the majority of observed torrents with a unique piece size, featured exactly 1088 pieces. The reason for this is unknown, as the bitTorrent standard gives no indication this should be occurring.

Finally, no **SHA-1** collisions occurred within the torrent set. The collision test was carried out on a Desktop computer with a 4.4 Ghz 4970k i7 Chip, the dataset was located on a **SSD** and the complete test took approximately 32 hours.



10.6 Evaluation of the BitTorrent SHA-1 Source

-ALEXANDER

The utilizing the torrent ecosystem as a **SHA-1** Rainbow table turns out to have some significant drawbacks.

The raw performance of the system is found by the **SHA-1** values obtained over time.

In other words, 3'287'001'641 **SHA-1** values divided by 2 weeks (1'209'600 seconds), which totals 2717,5 digests per second.

While this number is very low compared to the attainable throughput of a **GPU** or **CPU** brute-force solution, it is also primarily bounded by the magnet link resolution service. If the set of magnet links are already converted to torrent files the question is then, how does the set of torrent files perform as a **SHA-1** rainbow table.

To determine this, two concepts need to be covered. The Coverage and the Lookup speed.

The coverage can be extrapolated by assuming that the rest of the magnet links will be resolved at the same rate as the current set. I.e. a 40% resolution rate.

With this rate, the total set of torrents after a full run through of the magnet links is about 9 million individual files with a total of 12 billion digests.

While this may seem like an impressive number, compared to the digest space of **SHA-1**, it is only 8.210733×10^{-37} % coverage, which in turn means the probability of attaining any type of collision is very small.

Specifically, by using the numerical approximation from earlier the probability is:

$$1 - \frac{1}{e^{70312499994140625/1427247692705959881058285969449495136382746624}}$$

The speed when determining whether the rainbow table contains a key, is comparable to that of any other rainbow table implementation. Actually extracting the value is however potentially much slower and not guaranteed any success.

This is due to the fact that the torrent files themselves can point to a dead torrent swarm or the particular piece in question is hosted by a low bandwidth participant.

This might not have been the case when a torrent file was first recovered and added, but torrent swarms evolve and change over time and it would be infeasible for a rainbow table implementation to track each torrent swarm it contains a reference to.

In summary, utilizing the **BitTorrent** protocol as a source of **SHA-1** values is impractical with both respects to time and storage.

Furthermore, judging by the distribution of piece sizes and piece counts in torrent files. Even if the torrent network turned out to be an effective rainbow table, the distribution in the piece sizes between individual torrent files indicates that the protocol itself would not be any more susceptible to collision attacks than any protocol using **SHA-1**.

This is important since any colliding pieces found could be spread virally through the **DHT/PEX** mechanisms in order to stealthily corrupt torrent swarms, without any current mechanism to detect this. However, the protocol uses cryptographic hashing correctly and is therefore not susceptible to such an attack.



11 Alternative Attack Vectors

-LARS

While **SHA-1** is a core component of IT-security it is important to look at other factors as well.

Just as electricity, intruders follow the path of least resistance, this chapter will look at alternative attack vectors in order to quantify the proportional risk of the two types of **SHA-1** collisions compared to these existing threats.

11.1 Railway Methodologies

-LARS

To illustrate the risk assessment of the **E**uropean **R**ailway **A**gency, the following case was introduced with a request for **ERA** to reply with their assessment and reasoning, which is shown in chapter 12.2 Railway, page 80.

11.1.1 Safe Link Layer

-LARS

There are publicly available standards that governs wireless communication using **ERTMS**²¹¹, specifically: Subset 37²¹², 38²¹³, 57²¹⁴ and 92^{215, 216}.

One of the key components for integrity between the **STM** and **EVC** is the Safe link-Layer specified in subset 57. Leading to the following conundrum based on a study of the **ERTMS** standards²¹⁷:

STM FFFIS Safe Link Layer section 5.2.3.4 specifies that the authentication token is only 32 bits, with an unknown/unspecified algorithm²¹⁸ (**SECTION 5.2.3, 5.1.4**)

While **NIST** SP800-57 recommends -at least- 80 bits (in legacy mode) and 112+ bits.²¹⁹ (**PAGE 2**)²²⁰ (**TABLE 4, PAGE 67**)

Furthermore describing that truncated digests need to have an improved hashing algorithm.²²¹ (**PAGE 9-10**)

Expanding on that conundrum, the issue is that **NIST** recommends 112+ secure bits²²² while the **ERTMS** authentication message in safe-link only is 32 bits²²³ (section 5.2.3.4), with an unknown/ unspecified algorithm²²⁴ (5.2.3, 5.1.4) and that authentication is not needed for a "final disconnect" message (5.2.5.9), though the handling of such a message has been "*5.2.5.8.2 Intentionally deleted*".

The essential parts of the **NIST** recommendations related to this has been quoted below:

²¹¹ "Set of Specifications # 2 (ETCS Baseline 3 and GSM-R Baseline 0)."

²¹² "EuroRadio FIS - SUBSET-037."

²¹³ "Offline Key Management FIS - SUBSET-038."

²¹⁴ "STM FFFIS Safe Link Layer - SUBSET 057."

²¹⁵ "ERTMS EuroRadio Conformance Requirements - SUBSET-092-1."

²¹⁶ European Railway Agency, "ERTMS EuroRadio Test Cases Safety Layer - SUBSET-092-2," 092.

²¹⁷ "Set of Specifications # 2 (ETCS Baseline 3 and GSM-R Baseline 0)."

²¹⁸ "STM FFFIS Safe Link Layer - SUBSET 057."

²¹⁹ Barker and Roginsky, "Transitions."

²²⁰ Barker et al., "Recommendation for Key Management SP 800-57 Part 1: General Revision 3," 3.

²²¹ Quynh, "Recommendation for Applications Using Approved Hash Algorithms NIST SP 800-107 Rev. 1."

²²² Barker et al., "Recommendation for Key Management SP 800-57 Part 1: General Revision 3."

²²³ "STM FFFIS Safe Link Layer - SUBSET 057."

²²⁴ Ibid.



“For the Federal government, a minimum security strength of 112 bits is required for applying cryptographic protection (e.g., for encrypting or signing data). Note that prior to 2014, a security strength of 80 bits was approved for applying these protections, and the transitions in this document reflect this change to a strength of 112 bits

However, a large quantity of data was protected at the 80-bit security strength and may need to be processed (e.g., decrypted or have a digital signature verified).”

²²⁵Page 2

“5.1 Truncated Message Digest

Some applications may require a value that is shorter than the (full-length) message digest provided by an approved hash function as specified in FIPS 180-4. In such cases, it may be appropriate to use a subset of the bits produced by the hash function as the (shortened) message digest.

Let the (shortened) message digest be called a truncated message digest, and let λ be its desired length in bits. A truncated message digest may be used if the following requirements are met:

- 1. The length of the output block of the approved hash function to be used **shall** be greater than λ (i.e., $L > \lambda$).*
- 2. The λ left-most bits of the full-length message digest **shall** be selected as the truncated message digest.*

For example, if a truncated message digest of 96 bits is desired, the SHA-256 hash function could be used (e.g., because it is available to the application, and provides an output larger than 96 bits). The leftmost 96 bits of the 256-bit message digest generated by SHA-256 are selected as the truncated message digest, and the rightmost 160 bits of the message digest are discarded.

- 3. If collision resistance is required, λ **shall** be at least twice the required collision resistance strength s (in bits) for the truncated message digest (i.e., $\lambda \geq 2s$).*

These specifications for truncating the output of a cryptographic hash function promote application interoperability in situations where the use of shortened message digests is appropriate (and permissible), as determined by implementers and application developers acting in conformance with NIST Standards and Recommendations.

Truncating the message digest can impact the security of an application. By truncating a message digest, the expected collision resistance strength is reduced from $L/2$ to $\lambda/2$ (in bits). For the example in item 2 above, even though SHA-256 provides 128 bits of collision resistance, the collision resistance provided by the 96-bit truncated message digest is half the length of the truncated message digest, which is 48 bits, in this case.

The truncated message digest of λ bits provides an expected preimage resistance of λ bits, not L bits, regardless of the hash function used.

The expected second preimage resistance strength of a message digest truncated to λ bits sometimes depends on the length of the message. This dependence is determined as specified in Appendix A. Note that there are situations for which the expected second preimage resistance strength does not depend on the message length. For example, a 130-bit truncated message digest generated using SHA-256 has an expected second preimage strength of 130 bits, rather than a value in the range specified in Table 1 above for SHA-256. Truncating the message digest can have other impacts, as well. For example, applications that use a truncated message digest risk attacks based on confusion between different parties about the specific amount of truncation used, as well as the specific hash function that was used to produce the truncated message digest. Any application using a truncated message digest is responsible for ensuring that the truncation amount and the hash function used are known to all parties, with no chance of ambiguity.” ²²⁶Page 9-10

²²⁵ Barker and Roginsky, “Transitions.”

²²⁶ Quynh, “Recommendation for Applications Using Approved Hash Algorithms NIST SP 800-107 Rev. 1.”



It is comparable to a situation where too few bits for session identification in PHP prior to 2010.

11.1.2 Low Entropy Session Identification

-LARS

Before 2010, with PHP versions lower than 5.3.2 there were problems with the entropy of `session_start()` pseudo random data used for session ID cookies. While it was supposed to be 160 bits of data ensuring unique and un-guessable (random) data to confirm the identity of a user (authentication). It was not. Many parts could be deduced as it consisted of the following parameters:

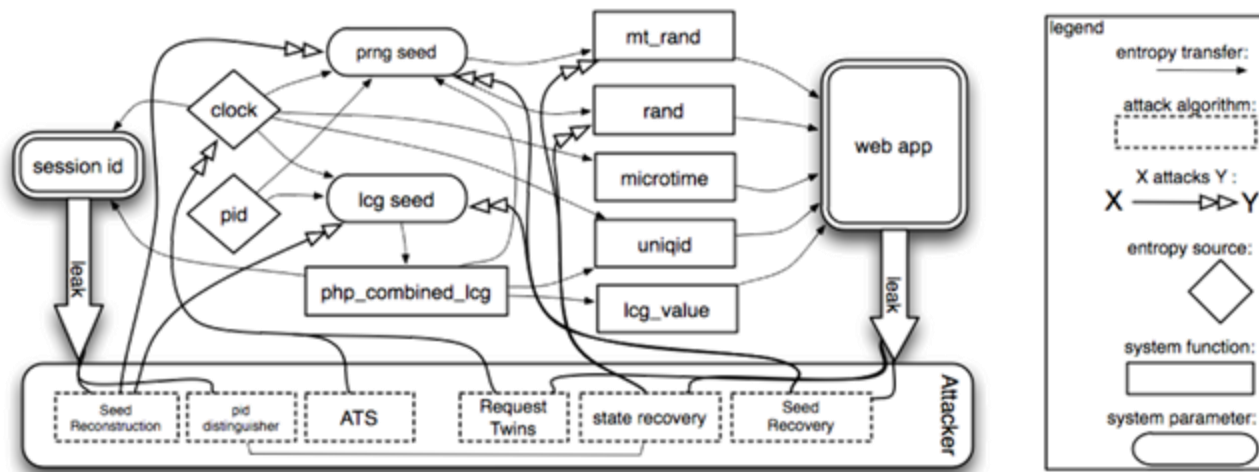


Figure 33 PRNG attack overview, by ²²⁷

- IP address: 32 bits
- Unix timestamp: 32 bits
- Millisecond: 32 bits
- Random `lcg_value()` (PRNG): 64 bits from two linear congruential generators
- Total: 160 bits

Milliseconds only have a sample space of 106 adding padding of 0 for 12 bits.

Unix timestamp: If the user logs on and has enabled chat the time stamp can be derived, decreasing the sample space one. Reducing the session id entropy by 32 bit.

IP address: If the user clicks a custom link to a server owned by the attacker the user's IP can be found in the access log. Reducing the session id entropy by 32 bit.

PRNG: The 64 bit seed consists of 2 parts each of 32 bits.

Part 1 is an XOR of Unix timestamp at server restart with milliseconds at server restart, meaning that if the restart time of the server is known within 12 days eg. by forcing a restart with flooding, entropy will be a reduced by 12 bits.

Part 2 is 32 bits for the process ID number, on Linux the process ID is only 15 bits long, giving a reduction of 17 bits. This can be further reduced if the attacker has access to `getmypid()` function via an error message making all of part 2 known.

Leaving **PRNG** (Part 1 + Part 2) to be $20 + 15 = 35$ bits or $20 + 0 = 20$ bits rather than 64 bits.

Giving a total reduction of 120 bits.

Resulting in a session ID of only 40 bits.

²²⁷ Argyros and Kiayias, "PRNG."



But as the 40 bits consists of two different 20 bit values (milliseconds & **PRNG**) they can be deduced individually. As **PRNG** is a known algorithm the valid seed bits can be guessed by brute force of the 20 bits on a local computer in a few seconds. Leaving only 20 bits, or an outcome space of 1'048'576 values.

This is a classic example of reliance on few bits of data for identification. One has to be very careful when assigning strong authentication through low number of bits as it lowers the resources needed for spoofing.

11.1.3 American Railway Risk Model

-LARS

As Europe has a harmonized set of signalling standards, it is interesting to compare it with the systems and standards in USA.

Currently, work is being done on a harmonized (federal) system: **Positive Train Control (PTC)**²²⁸. Research into a "Composite risk model for railroad operations utilizing Positive Train Control"²²⁹ centered around IT-security has been done by Wijesekera Duminda, whom have been unavailable for comments, leaving **PTC** security as future work.

11.1.4 Open ETCS

-LARS

In the efforts made to produce this report, while some standards required a fee to be read and other were freely available there has been a general problem with openness:

Even when the text is available, test systems are not, nor proper structure or references.

Leading to a system that is more likely to have latent and persistent vulnerabilities as they are hard to find.

There is a proposal for making the **ETCS** system not only **Open Source Software (OSS)**, but *open proof*²³⁰, a term suggested by the US military think tank "Institute for Defense Analysis"; that not only the software itself, but any tools used in the validation process should be *open source*, extending as far as requiring open training material and documentation enabling anyone interested to test the system²³¹.

Sending specific questions to relevant authorities has on several occasions lead to the answer of:

"All the specifications can be downloaded at: <http://www.era.europa.eu/Core-Activities/ERTMS/Pages/Set-of-specifications-2.aspx>"²³²

Indeed, they can, but there are 80 documents with several of them spanning hundred pages in a highly technical language.

In an environment where answers are along the lines of -it is probably somewhere within these thousands of pages-, not even referencing a volume, chapter or page it is near impossible to make suggestions for improvements or test security.

Referencing that an answer would be somewhere in all the standards is safe from a management perspective: It places responsibility on the standard, relying on it to be secured by its authors.

²²⁸ Joint Council on Transit Wireless Communications, "Positive Train Control White Paper."

²²⁹ Abadie, Bandara, and Wijesekera, "A Composite Risk Model for Railroad Operations Utilizing Positive Train Control (PTC)."

²³⁰ Hase, "'Open Proof' for Railway Safety Software - A Potential Way-Out of Vendor Lock-in Advancing to Standardization, Transparency, and Software Security."

²³¹ Institute for Defense Analyses, "Open Source Software (OSS/FLOSS) and Security International Workshop on Free/Open Source Software Technologies Riyadh, Saudi Arabia."

²³² European Railway Agency Corporate Management and Evaluation, "FW: Information Request Form - Nielsen (Dec 2)," 2.



Exploring the standard, finding an issue, a responsibility to act upon it arises, an effort requiring a highly specialized skill set and money to verify.

If the issue suddenly needs fixing, it is likely to cost a lot of money when it is already implemented.

Engaging in a conversation on why something is not safety critical would lead to many more bugs and issues being discovered, as illustrated in the theory section of this report; chapter 2.2.1.2 Responsible Disclosure in a Risk Assessment Perspective, page 18.

But it would also incur a short term loss from the resources used to manage and verify the input.

The UNISIG standards are already sent out to be peer reviewed by private sector actors in the domain making it a living standard, with various baselines. But it lacks a channel for public feedback, and a way to incentivize independent security researchers to provide feedback and better the next generation of railway standards.

11.2 NemID

-LARS

As part of the Danish government's effort to enable web authentication, digital signatures etc. the product "Secure E-mail" is provided by the Danish Agency for Digitisation (Digitaliseringsstyrelsen).

This signing service and the 2-factor authentication scheme "NemID" it uses is designed by the company "nets". It uses a X.509 infrastructure based on a central public CA (Certificate Authority). (In Danish the word "Public" and "Government" is the same word, leading to some confusion)

Certificate signing is using **SHA-256**, even for fingerprints, specified and published in the Danish Trusted Service List.²³³

There is a claim that the Trusted Service List is a requirement by "Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on a Community framework for electronic signatures", but the word "list" does not appear in that directive and article 3 only details that:

"Member States may introduce or maintain voluntary accreditation schemes aiming at enhanced levels of certification-service provision. All conditions related to such schemes must be objective, transparent, proportionate and non-discriminatory. Member States may not limit the number of accredited certification-service-providers for reasons which fall within the scope of this Directive"

Leading to no requirement on a Trusted Service List in the referenced directive, but only a framework for voluntary accreditation schemes. Furthermore the primary focus on those rules are qualified certificates.²³⁴

The Danish national authentication service (NemID) is explicitly not a qualified certificate and hence not covered by the requirements, specifically "*Requirements for certification-service-providers issuing qualified certificates*" in ANNEX II of the directive: "(j) not store or copy signature-creation data of the person to whom the certification-service-provider provided key management services;", ratified in the Danish law as Act no. 417 of 31 May 2000 on Electronic Signatures.^{235, 236}

The requirements are instead specified in the "*Certificate policy for OCES (Public Certificates for Electronic Services)*" that explicitly details how it is not detailing qualified certificates, but a ruleset that is less strict.^{237, 238}

²³³ "Trusted Service List - Dansk."

²³⁴ Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on a Community Framework for Electronic Signatures, 6.

²³⁵ "DANMARK (DENMARK) : Trusted List."

²³⁶ Lov Om Elektroniske Signaturer (Act No. 417 of 31 May 2000 on Electronic Signatures).

²³⁷ Danish Agency for Digitisation and Triantafyllidis, "Certifikatpolitik for OCES-Personcertifikater (Offentlige Certifikater Til Elektronisk Service) Version 4."

²³⁸ Danish Agency for Digitisation, "Certificate Policy for OCES Employee Certificates (Public Certificates for Electronic Services)."



Some of the reasoning for the **OCES** relaxed ruleset is specified as:

"In addition, qualified certificates exist that have been issued in pursuance of Act no. 417 of 31 May 2000 on Electronic Signatures. A qualified certificate is not based on the above-mentioned common public standard. Among other things, personal attendance is required when issuing a qualified certificate."

It has not been possible to verify this requirement of personal attendance. It is not present in either the *"Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on a Community framework for electronic signatures"* or the ratified Danish equivalent *"Act no. 417 of 31 May 2000 on Electronic Signatures"*.

ANNEX II, Requirements for certification-service-providers issuing qualified certificates & § 6 in the Danish ratification only detail that:

*"Certification-service-providers must:
(d) verify, by appropriate means in accordance with national law, the identity and, if applicable, any specific attributes of the person to which a qualified certificate is issued;"*

	Directive 1999/93/EC	Act no. 417 of 31 May 2000; DK ratification of Directive 1999/93/EC:	Certificate policy for OCES (Public Certificates for Electronic Services)	ETSI TS 102 231 V3.1.2 (2009-12)
Mentions qualified certificates need physical presence	No	Yes	Yes	Yes
"Public" is replaced with "Government"	No	Yes	Yes	No
Mentions TSL	No	No	Yes	Yes
Mentions TSL as a requirement	No	No	No	Yes

Additionally, the Danish word for "Public" is the same as the one for "Government", leading to some confusion around the **OCES** name akin to the confusion about "free" meaning both "liberty" and "gratis".

"Thus, the basic principle governing the CP[Certificate Policy] is that the public authority that holds the main responsibility for the field in question, i.e. the National IT and Telecom Agency, prepares it."²³⁹

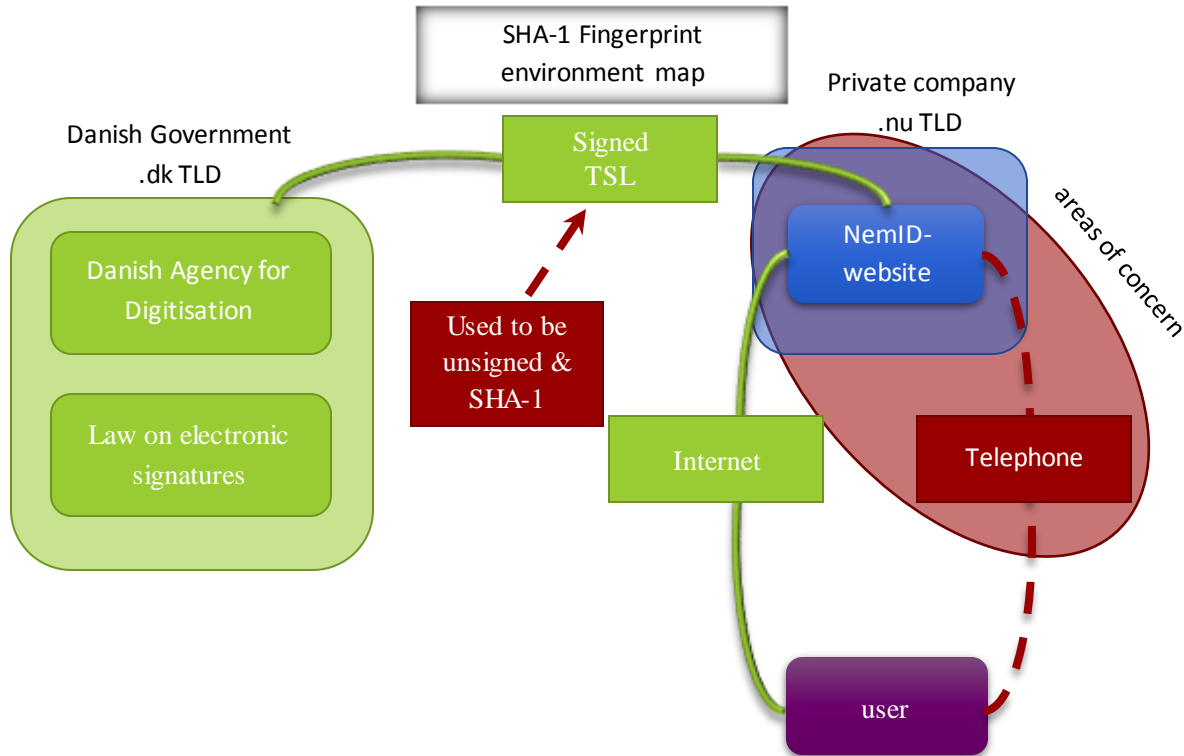
"The National IT and Telecom Agency is the public authority which authorises the issue of OCES employee certificates for the selected certification authorities (CAs), and which is in charge of the approval of the CAs in accordance with this CP[Certificate Policy]."²⁴⁰

These two quotes illustrate the problem of the Danish translation, where "government authority" in the English translation of the Danish text is written as "public authority", leading to the conclusion that Danish certificates are government issued and government backed, rather than being open and public as the EU directive specifies.

It is specified that the Danish **OCES** certificate indeed not is qualified, but it goes against the intention of the EU directive yet it is a requirement for Danish citizen to use for municipal or government contact and interaction.

²³⁹ Ibid.

²⁴⁰ Ibid.




We have reached out to **ENISA** several times, asking for documents that has been signed with this **SHA-1** certificate, but **ENISA** has not replied to our inquiries.

11.2.1 SHA-1 Root Certificate Verification

-LARS

Recalling the certificate users are shown, displayed page 45 Figure 22 "User experience for "Secure mail" - showing the name of TRUST2408 OCES" showing a **SHA-1** value for the "TRUST2408 OCES Primary CA". Going to the website named "rules" in Danish²⁴¹ the translated text says:



Verification of the root certificate by telephone. The service provider can verify whether a root certificate is correct by calling +45 80 30 70 12. By comparing the fingerprint found in the root certificate with the fingerprint read aloud on the telephone, the correctness of the root certificate can be confirmed.

Figure 34 Page 33 in Implementation guidelines for NemID (OCES)²⁴²

First of all the user is only shown a **SHA-1**, secondly the phone number has not been working from at least November 15th to November 26th going well into December, where an automatic voice replied with: "The dialled number does not exist" in both English and Danish.

Trying to find out if we could purchase that phone number and own the phone line advertised as being the root verification for the Danish national digital ID service we found out that it was in a range of numbers they only sold to companies.

So we spent 100 **USD** and made a company in Denmark in 7 hours (HPC Frontrunners IVS, CVR 37244767) to get hold on the list of available phone numbers starting with 80 30 and ending in 0 12.

The list can be seen in Table 3.

²⁴¹ "Regler - Om NemID - NemID (verified January 11-2016)."

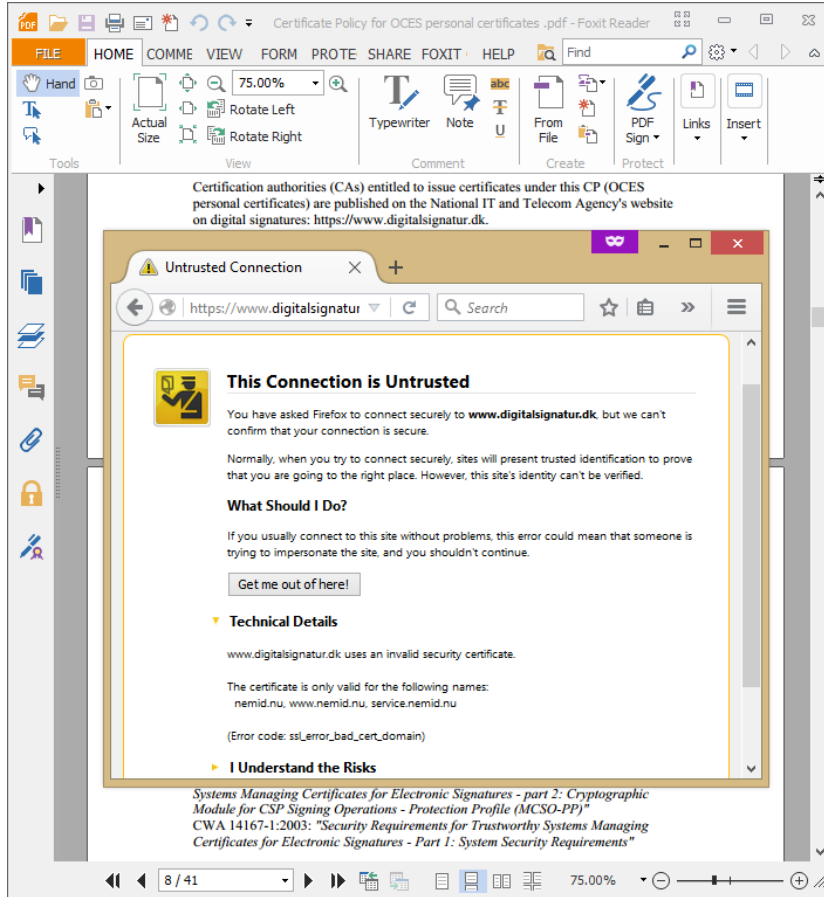
²⁴² "Implementation Guidelines for NemID (OCES) Version 2.1."



Sadly the targeted number was not available, but if it does become available one day, it can be purchased for ~300 USD and a quarterly fee of ~100 USD.

On a check-up January 11th 2016 the phone number was found to work again, also offering a **SHA-256** digest.

A second discrepancy can be seen below:



Available phone numbers
80 30 10 12
80 30 20 12
80 30 30 12
80 30 40 12
80 30 50 12
80 30 60 12
80 30 80 12
80 30 90 12

Table 3 Company phone numbers available 26/11-2015

Sadly the targeted phone number was not available

Figure 35 Certificate Policy for OCES personal certificates (Public Certificates for Electronic Services) in the background and in the foreground the URL specified in the policy²⁴³.

Where the certificate policy specifies the place look up the list of government verified Certificate Authorities. This website redirects to www.nemid.nu a domain outside of the Danish .dk domain, the island state of Niue with a GDP of 10 million USD. It is the official website of the currently only OCES Certificate Authority though, but having invalid certificates and redirecting users away from the national Top Level Domain is normally a sign of phishing.

We have reached out to Nets as well, which resulted in some good initial contact, but we have been unable to reach them for comments in the last months even when including some of the discrepancies mentioned above. The phone number 80 30 70 12 does seem to work for root certificate verification now though.

²⁴³ Danish Agency for Digitisation, "Certificate Policy for OCES Personal Certificates (Public Certificates for Electronic Services)."



12 Impact Analysis

-LARS

This chapter explores the impacts of some existing IT catastrophes to find the monetary loss for these instances, in order to estimate a cost for **SHA-1** attacks.

While **SHA-1** is widely in use as shown in chapter 5 Current use of SHA-1, pages 34-42, there is also a movement towards newer and safer hashing algorithms. The move is primarily driven by big software companies such as Google and Microsoft having announced January 2016 as the deprecation date for **SHA-1** in their products.²⁴⁷,²⁴⁸

While this change was announced by Microsoft November 2013 and Google August 2014 the move away from **SHA-1** has been slow.

The distinction between price and cost is imperative; price being the money spent on an attack, while cost is the loss the attack incurs.

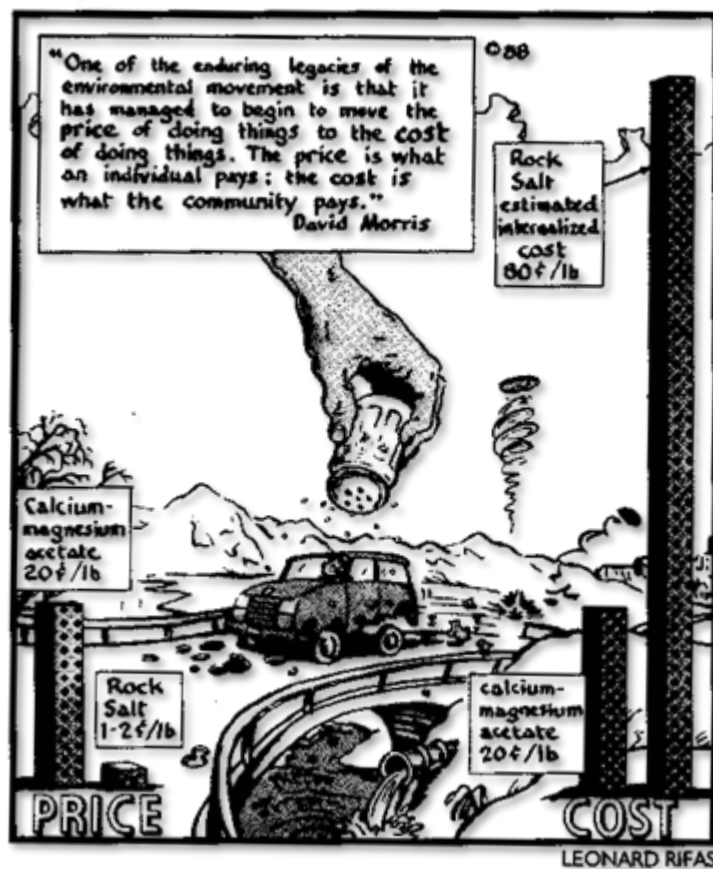


Figure 36 Distinction between Price (what is paid) and Cost (decrease in profit it causes)
The illustration above shows this difference and this definition will be used throughout this chapter.

²⁴⁷ "Intent to Deprecate: SHA-1 Certificates - Google Groups."

²⁴⁸ "SHA1 Deprecation Policy - Windows PKI Blog - Site Home - TechNet Blogs."



12.1 Denial of Service

-LARS

A **D**enial **O**f **S**ervice (**DOS**) attack is one that hinders availability, it is any attack that makes a website, server or service unavailable, this sub-chapter will detail some documented impacts of this in order to argue a reference in known attacks for *Risk Class* and a recommended **SIL** level.

Where a **DOS** targeted towards a single company of cause is of lesser severity than a **SHA-1** attack that has a much broader attacksurface.

In December 2010 Paypal (owned by ebay) was hit by a **DOS** attack for 10 days:

"More than 100 workers from PayPal's parent company, eBay, spent three weeks working on issues related to the attacks"

"PayPal also had to pay for more software and hardware to defend against similar attacks in the future ... the total cost to the firm was estimated at £3.5m"

-Sandip Patel, Prosecutor in UK 2012 case on **DOS**ing paypal and ebay²⁴⁹

Which was **USD** 5,5 m | **€**4,1 m | **DKK** 3.8 m with a December 2010 average conversion rate, Estimated costs for the **DDOS** attack in 2010²⁵⁰:

The Ministry of Sound: £9'000

International Federation of the Phonographic Industry: > £20'000

British Phonographic Industry: > £4'000

By adjusting those court case numbers for inflation an upper bound for the cost of a **DOS** attack can be established and hence a meter for the availability metric.

Combined with the **ALARP** from annex C of 61508 and Value of Preventing a Fatality an appropriate **SIL** can be found.

The reason this is an upper bound estimation is because it includes rush fees for consultants and overpay for long work hours, fees that would not be necessary if the work had been pre-emptive, on the other hand it does not cover the costs of the individual users of the system losing access to PayPal for the 10 days it lasted, meaning this is not the cost for society, but the individual companies, to find the damage to society those externalities should be accounted for.

²⁴⁹ "Anonymous Hackers 'Cost PayPal £3.5m.'"

²⁵⁰ "Anonymous Hacker Group."

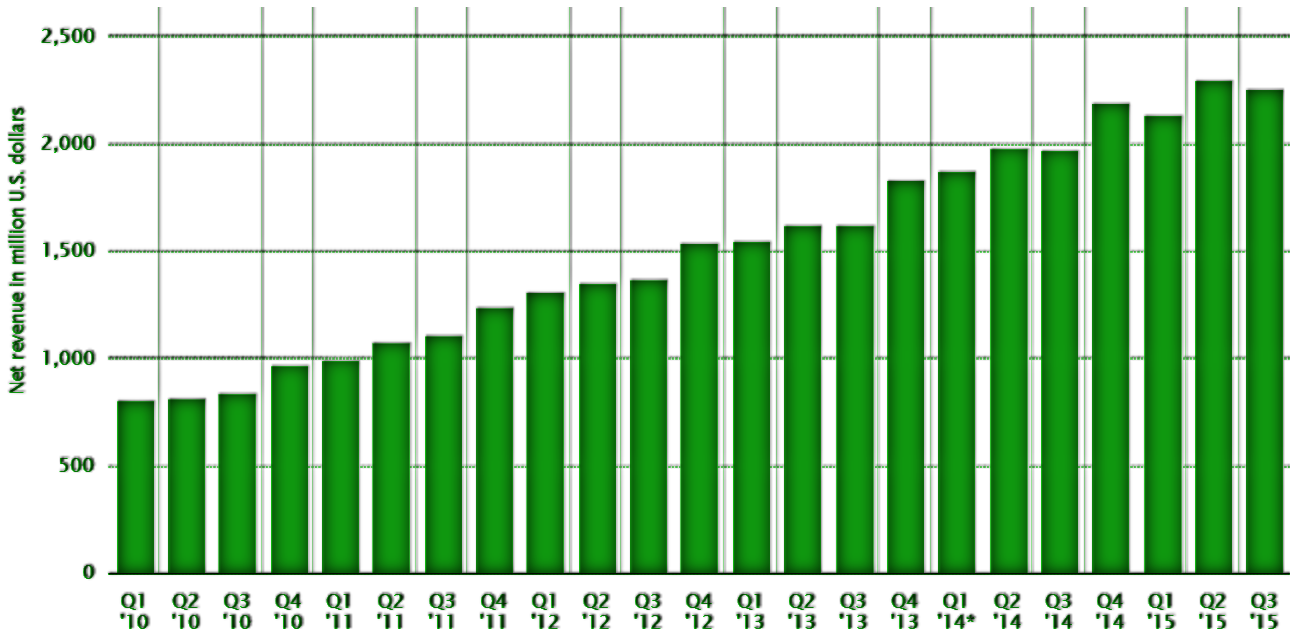


Figure 37 Paypal quarterly revenue²⁵¹

While the attack took place in Q4 2010, it did not seem to influence the revenue significantly, compared to later years the difference between Q4 and Q1 are similar. The yearly net revenue for Paypal 2010 was 3'435 million USD | £2'224 m | € 2'594 m | DKK 19'320 m with a 2010 average conversion rate, leading the Paypal DOS attack to account for 1,5‰ of the yearly revenue, yet 10 days account for 27‰ of a year.

With the data at hand it is hard to spot a significant impact of the DOS attack on the yearly or Q4 revenue 2010.

Paypal reported²⁵² that a large amount of the estimated cost was for wages and hardware associated with future mitigation of DOS, leading to the assumption of some economy of scale, meaning that the mitigation measures most likely would take a larger size of the revenue of smaller companies and that all revenue would be lost during the DOS, rather than just 0,15‰ per day, it would be 2,7‰

Translating the DOS attack cost of €4,1 m to the Danish Value of a Statistical Life in 2010²⁵³ (€2'724'418.60465) it translates to 1½ fatalities.

“Frequent could denote an event that is likely to be continually experienced, which could be specified as a frequency greater than 10 per year. A critical consequence could be a single death and/or multiple severe injuries”

from ²⁵⁴(CHAPTER 5), based on ²⁵⁵.

Leading to Risk Class I, requiring SIL 4.

²⁵¹ “Paypal.”

²⁵² “Anonymous Hackers ‘Cost PayPal £3.5m,’” 5.

²⁵³ “Common Safety Indicators, Denmark 2010, Version 5, Validated.”

²⁵⁴ International Electrotechnical Commission, “IEC 61508 Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems.”

²⁵⁵ Great Britain. Health and Safety Executive, *Reducing Risks, Protecting People*.



12.2 Railway

-LARS

Investigating the Safe Link Layer in chapter 11.1.1, pages 69 to 72 lead to the investigation of a set of previous accidents that illustrate why message integrity and authentication is important.

The following inquiry was sent to **ERA**, and finally **ENISA**:

STM FFFIS Safe Link Layer section 5.2.3.4 specifying that the authentication token is only 32 bits, with an unknown/unspecified algorithm²⁵⁶ (SECTION 5.2.3, 5.1.4)

While **NIST** SP800-57 recommends -at least- 80 bits (in legacy mode) and 112+ bits.²⁵⁷ (PAGE 2)²⁵⁸ (TABLE 4, PAGE 67)

Furthermore describing that truncated digests need to have an improved hashing algorithm.²⁵⁹ (PAGE 9-10)

Inquiry sent to **ERA** based on a study of the **ERTMS** standards²⁶⁰.

The answer from **ERA** came in two parts:

- 1) Pointing out that IT security is not the responsibility of **ERA**, but **ENISA**
- 2) Stating that a masquerade attack would need²⁶¹:
 - a. -to get physical access to the cab train,
 - b. -to be able to power up a train,
 - c. -to introduce the correct parameters for a train mission,
 - d. -to hack the interface,
 - e. -to provide correct signalling information.

(Full letter can be found in Appendix, 20.5 ERA letters, page 126)

By showing methods that circumvent the points of defence, it is possible to substantiate an impact.

These barriers listed are predominantly physical;

relying on restricted access to the train for security, putting it in category 2 "Category 2 consists of systems which are partly unknown or not fixed, however unauthorised access can be excluded" under EN 50109²⁶², making it imperative that vendors do not implement cables accessible by passengers to rely on the Safe Link Layer authentication message.

Specifying to **ERA** that a hypothetical attack could be:

"A remotely executed attack during regular operation that could eg. increase the allowed speed, leading to a derailment at a switch/turnout or curve. That is if the security relied on the Safe Link Layer the 4 byte authentication message."

Follow-up question to **ERA**²⁶³

Suggesting a use of the Safe Link Layer protocol in a category 3 environment "Category 3 consists of systems which are not under the control of the designer, and where unauthorised access has to be considered" under EC 50109²⁶⁴, as opposed to relying on physical barriers to hinder tampering.

²⁵⁶ "STM FFFIS Safe Link Layer - SUBSET 057."

²⁵⁷ Barker and Roginsky, "Transitions."

²⁵⁸ Barker et al., "Recommendation for Key Management SP 800-57 Part 1: General Revision 3," 3.

²⁵⁹ Quynh, "Recommendation for Applications Using Approved Hash Algorithms NIST SP 800-107 Rev. 1."

²⁶⁰ "Set of Specifications # 2 (ETCS Baseline 3 and GSM-R Baseline0)."

²⁶¹ European Railway Agency Corporate Management and Evaluation, "FW: Information Request Form - Nielsen (Dec 2)."

²⁶² "Railway Applications - Communication, Signalling and Processing Systems - Safety-Related Communication in Transmission Systems - EN 50159."

²⁶³ European Railway Agency Corporate Management and Evaluation, "FW: Information Request Form - Nielsen (Dec 3)."

²⁶⁴ "Railway Applications - Communication, Signalling and Processing Systems - Safety-Related Communication in Transmission Systems - EN 50159."



ERA supplied the following answer to this second scenario:

“1) the ERTMS is not an ATO system i.e. it is a protection system with a driver presence, I mean it is the driver who is driving not the ERTMS system. So, it looks that you would need some cooperation from the driver who needs route knowledge and speed tables to be allowed to drive.

2) Your “fake allowed speed” should come either from and RBC or a balise, so you should know the RBC and balise identifiers and get access to railway installations again.

Please bear in mind that if needed I could even change the keys every time I communicate, so that if you sniff the info it will not be usable for the next communication.

Our specifications does not mention when each key can be changed, it provides the mean to change it. It is up to each administration to do decide when, how often, ...

You could argue that the machine providing the keys can be hacked, of course yes as any IT system, but these machines are normally certified for security and this is beyond the ERTMS and ERA scope of work.”

Answer from ERA on remotely executed masquerade attack raising the maximum speed allowed²⁶⁵.

The listed barriers can be circumvented as follows:

1) **ATO**, meaning **A**utomatic **T**rain **O**peration is not the goal of **ERTMS**, the goal is safer, faster, more compact use of trains on the railway. **ERTMS** level 2+ (currently under implementation nationwide in Denmark) will also remove all trackside physical signals, so the driver relies 100% on the displays in the cabin, with information streaming from the Radio Block Center and balises²⁶⁶.

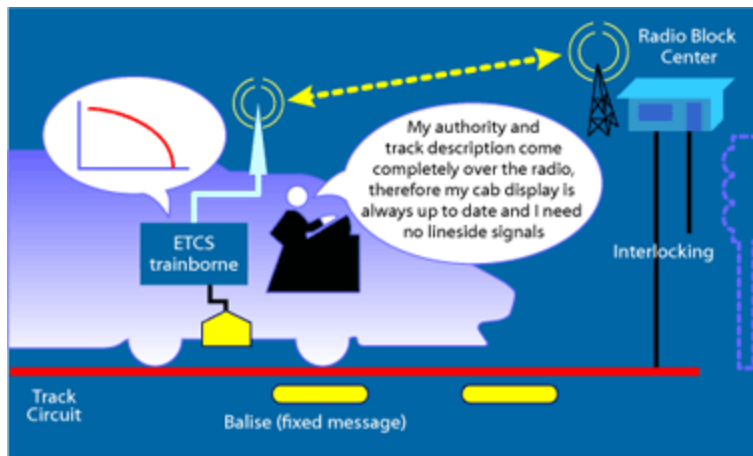


Figure 38 ERTMS level 2 diagram ©ERTMS.net

Secondly the Frutigen derailment October 16th 2007 is an example of an **ERTMS** software bug causing a derailment²⁶⁷ (German),²⁶⁸ (English summary).

²⁶⁵ European Railway Agency Corporate Management and Evaluation, “FW: Information Request Form - Nielsen (Dec 3).”

²⁶⁶ “ERTMS Signaling Levels | ERTMS.”

²⁶⁷ Schweizerische Eidgenossenschaft, “Frutigen ERTMS derailment report (Schlussbericht der Unfalluntersuchungsstelle Bahnen und Schiffe über die Entgleisung von Güterzug 43647 der BLS AG auf der Weiche 34 (Einfahrt Lötschberg-Basisstrecke) vom Dienstag, 16. Oktober 2007 in Frutigen).”

²⁶⁸ “ETCS Software Error Led to Lötschberg Derailment.”

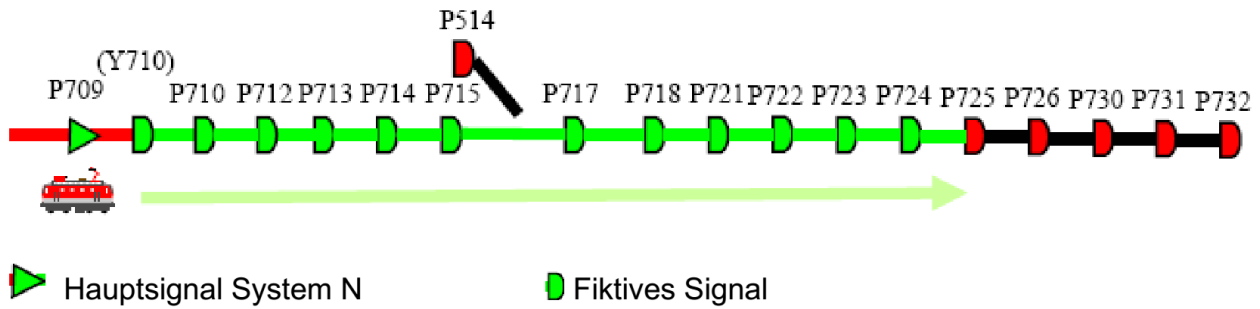


Figure 39 Physical main signalling system and virtual signals at Frutigen: Damages for 90+360 K€ from ²⁶⁹

Increased speed does cause derailments, even with drivers present, as illustrated by the Santiago de Compostela derailment in Spain 2013.



Figure 40 Santiago de Compostela derailment in Spain July 24th 2013. 79 dead, 140 injured
 The conclusion from this derailment in Spain was to incorporate automatic braking systems that would avoid derailment accidents based on speed even with driver error by “installation of three balises on 1.9 km of the approach to Santiago to enforce speed limits of 160, 60 and 30 km/h”²⁷⁰.

A well-known aviation case defining the regulation of trust in technical aids is the Überlingen mid-air collision (69 dead) where the **Traffic Collision Avoidance System (TCAS)** was ignored by the flight controller and pilots, leading to regulations sanctioning tighter reliance on automated computer systems, declaring **TCAS** to have authority above that of the flight controller:

“Pilots flying are required to obey and follow TCAS resolution advisories (RAs), regardless of whether contrary ATC instruction is given prior to, during, or after the RAs are issued.”
 Safety Recommendation No. 18/2002, ²⁷¹

While a good train driver should know the track and the speeds for safe travel, the cases above shows an increased reliance on automated systems to tell the truth and have better judgements than human operators. While safety has the highest priority, a driver seeing a higher allowed maximum speed is encouraged to utilize the speed in a way that will give the least transportation time. Making masquerade attacks more likely to have an impact.

²⁶⁹ Schweizerische Eidgenossenschaft, “Frutigen ERTMS derailment report (Schlussbericht der Unfalluntersuchungsstelle Bahnen und Schiffe über die Entgleisung von Güterzug 43647 der BLS AG auf der Weiche 34 (Einfahrt Lötschberg-Basisstrecke) vom Dienstag, 16. Oktober 2007 in Frutigen).”

²⁷⁰ “Further Safety Measures Follow Santiago de Compostela Crash.”

²⁷¹ German Federal Bureau of Aircraft Accidents Investigation, “Überlingen Mid-Air Collision Investigation Report.”



Figure 41 Bombardier ERTMS Level 2 High Speed Eurobalise © Bombardier, from press release²⁷²

2) Eurobalises are placed in open land in remote areas, getting access to them, the information and their identifiers is not a problem²⁷³.

The third argument that *“if needed I could even change the keys everytime I communicate”* is hard to counter as there is no indication of who or what “I” covers in that sentence. While it was sent from an official ERA address, there was no name given and we were referred to ENISA for further inquiries.

Combined with the claim of *“Our specifications does not mention when each key can be changed, it provides the mean to change it. It is up to each administration to do decide when, how often, ...”* it hints to be either the symmetric encryption keys mentioned in subset 38²⁷⁴ or the three triple DES keys used for message authentication in Euro Radio FIS mentioned in subset 37²⁷⁵.

None the less, it does not alter that the Secure Safety Layer uses 32 bits to authenticate messages, a choice that seems strange in relation to the use of 191bit keys (112 secured bits²⁷⁶) for 64bit MACs and NIST recommendations.

Designing an IT system for the future, expecting at least 14 years of usage, more likely going for 30 to 40 years, relying for 32bit authentication codes seems to be an inefficient place to save money, given the high cost of the physical installations, a 192bit (24 byte) digest does not seem unreasonable. Even if time was the issue, a change from triple DES to AES would save time and 1 second response time is tolerable, up to 5 seconds before it has a safety impact²⁷⁷ (SAFEDMI REQ 7).

Referring to chapter 11.1.2 Low Entropy Session Identification page 71.

²⁷² “Bombardier Enters ERTMS Level 2 High Speed Rail Control Market in Spain - Bombardier.”

²⁷³ “ERTMS Signaling Levels | ERTMS.”

²⁷⁴ “OfflineKey Management FIS - SUBSET-038.”

²⁷⁵ “EuroRadio FIS - SUBSET-037.”

²⁷⁶ “Expert Advice.”

²⁷⁷ Jørgensen, “Analysis and Enhancement of Safety Critical Communication for Railway Systems.”



Throughout this chapter it has been illustrated that:

1. ERTMS can have derailments due to software bugs (Frutigen)
2. Trains derail when driving too fast (Santiago de Compostela)
3. There is open access to trackside equipment
4. There is an increased reliance on automated system data (aeronautics)
5. 32 bits of entropy is too little (PHP PRNG/**NIST**)

Making masquerade attack quite plausible and can be used to illustrate the size of economic impact derailments have.

As with the previous chapter this leads to a *Risk Class* of I, requiring **SIL 4**.

12.3 Heartbleed

-LARS

The **OpenSSL Heartbleed** bug is in many ways comparable to a **SHA-1** exploit: It relies on a security feature that is in widespread use and embedded in many security systems.

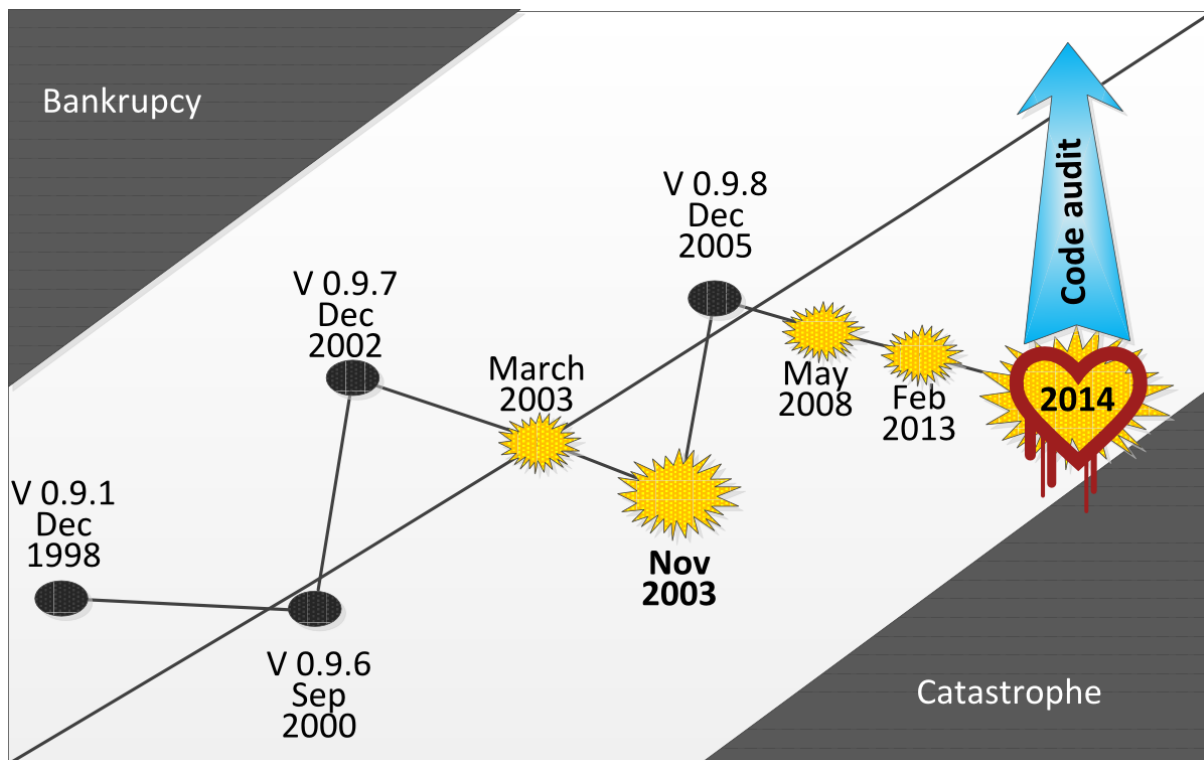


Figure 42 Rocking boat, by Lars Embøll, derivative of ²⁷⁸

Heartbleed is an example of the complacency effects described by the bathtub and rocking boat principle (chapter 2.1.1, page 15); reliance on old trusted code, trust that an open source format ensures security by transparency. It was only after an exploit that the **OpenSSL** community got funding for a thorough investigation²⁷⁹ as the whole world had already implemented **OpenSSL** in a vast amount of security installations.

While other security incidents have rocked the boat of **OpenSSL**, such as the 2008 Debian specific incident and incremental strengthening with updates and newer versions released, there have been a widespread and increased reliance on an aged system based on trust rather than checks as the 2013 timing attack and 2014 **Heartbleed** incidents show.

²⁷⁸ Reason, *Managing the Risks of Organizational Accidents*.

²⁷⁹ "OpenSSL Audit."



But what makes Heartbleed especially interesting is the immediate response and mitigation:

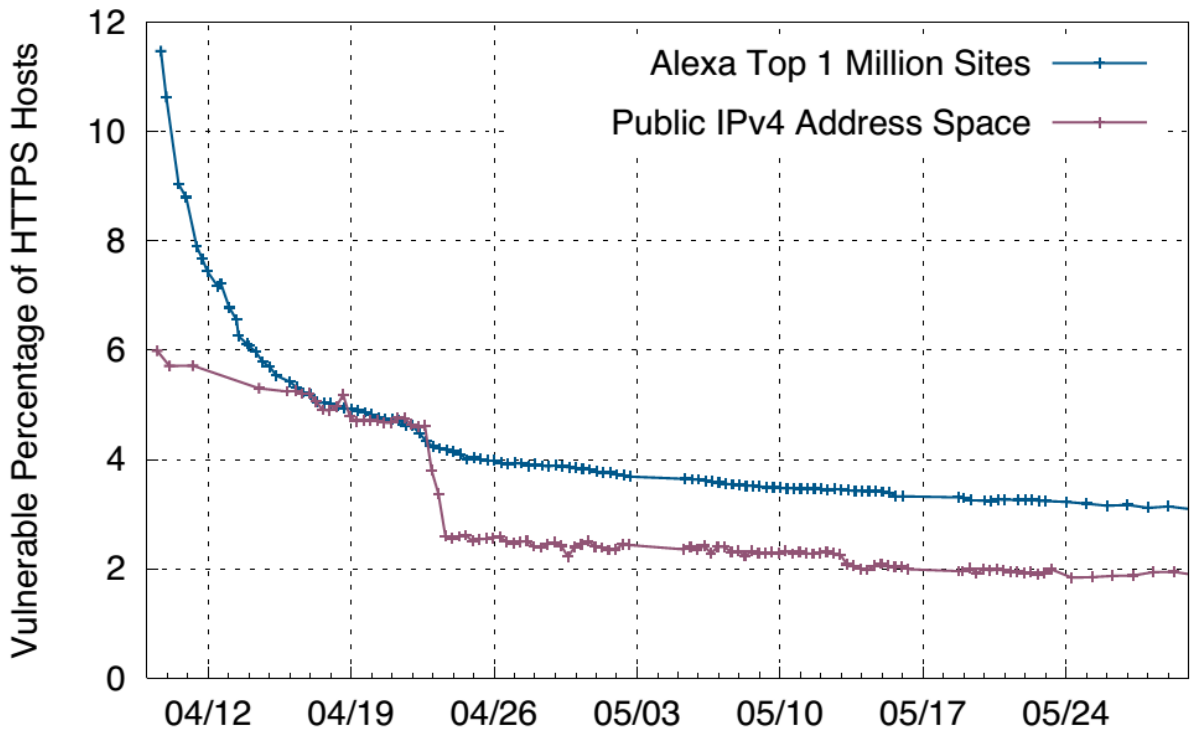


Figure 43 Patch propagation after Heartbleed publication, by ²⁸⁰

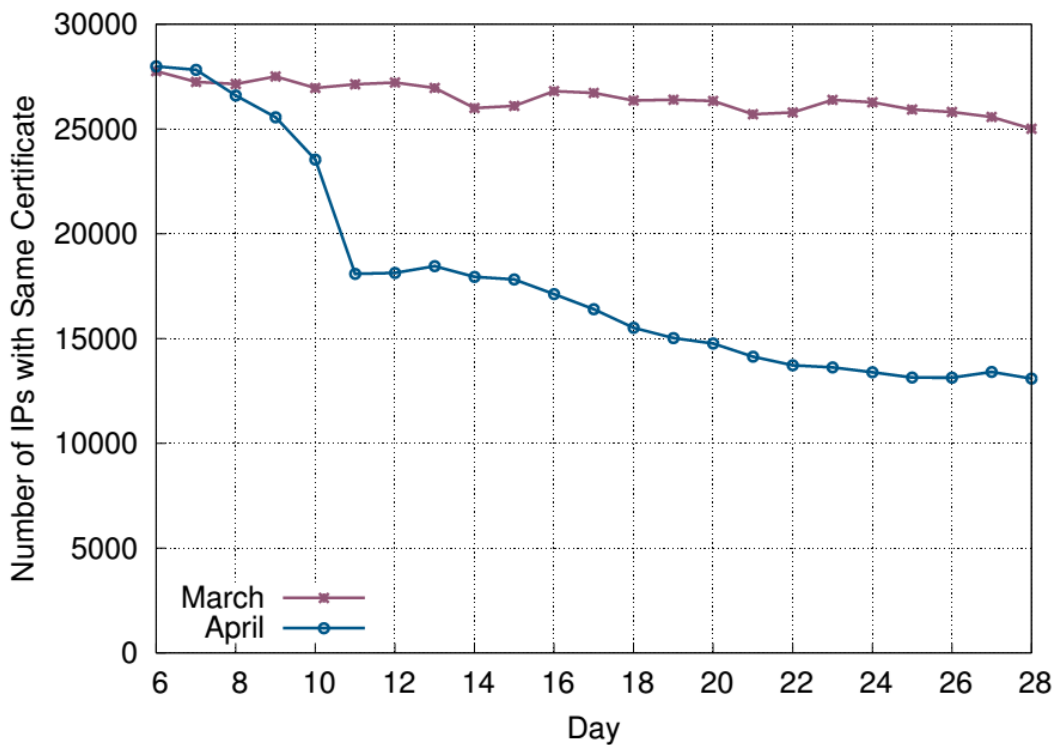


Figure 44 Website certificate changes before and after Heartbleed publication, by ²⁸¹

²⁸⁰ Durumeric et al., "The Matter of Heartbleed."

²⁸¹ Ibid.

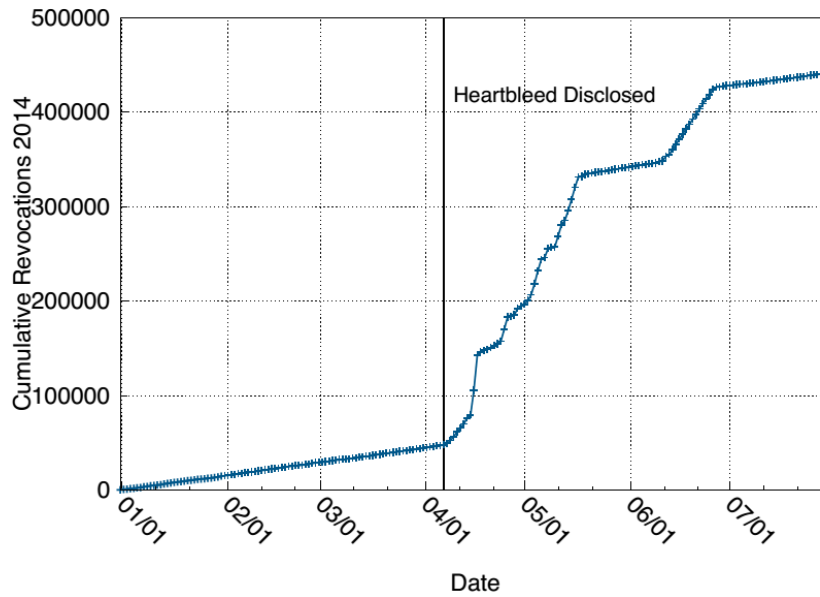


Figure 45 Website certificate revocations after Heartbleed, by ²⁸²

The big jumps in the graph just above are when GlobalSign first and GoDaddy secondly revoked their certificates. It is also worth noting that in the Alexa top 1 million sites using **SSL/TLS** only 4% revoked their certificates between April 9 and April 30 2014.

This response was costly though; GlobalSign was reported to spend more than 400'000 **USD** per month to handle the revocation of certificates due to **Heartbleed**.²⁸³

This response also marks the difference between **Heartbleed** and a **SHA-1** exploit:

There is no revocation for **SHA-1** as it is embedded on a much deeper and fundamental level than **SSL/TLS**. One does not simply replace **SHA-1**, as it often is an embedded chip in hardware, as well as a fundamental security feature described as mandatory in some standards, as described in chapter 5.5 openPGP, page 40 leading to possible downgrade attacks as long as **SHA-1** is a mandatory option.

12.4 Chapter Summary

-LARS

Prior incidents show a pattern of likelihood and impact for **SHA-1** masquerade attacks, this chapter will sum it up in a table with the same currency, adjusted by the average European price index²⁸⁴ to be in 2014 prices:

Table 4 Impacts of previous comparable incidents, in €, 2014 prices

Incident	Impact
<i>Heartbleed</i>	>289'300€ / month (single company, March 2014 average exchange rate)
<i>Derailment</i>	103'846 +415'385 € (Frutigen, 2007 prices) 79 dead & 140 injured (Santiago de Compostela)
<i>DOS</i>	(December 2010 exchange conversion) 4'262'730 € PayPal 11'042 € The Ministry of Sound 4'906 € British Phonographic Industry 24'526 € International Federation of the Phonographic Industry

²⁸² Ibid.

²⁸³ "The Hidden Costs of Heartbleed."

²⁸⁴ "Open Data Catalog | The World Bank."



With these specific impacts in place, the broader consequences can be examined in the next chapter.

13 Consequence Analysis

-LARS

While the previous chapter described the price of individual impacts the overall consequences and costs for society in the event of a **SHA-1** 2nd pre-image vulnerability relies on estimation based upon those previous cases.

It should be noted that the estimates in this chapter are conservative, trying to establish a lower bound.

If this lower bound requires precautionary measures of **SIL** 4 for a **SHA-1** 2nd pre-image vulnerability it is hence documented that no less than **SIL** 4 is required.

With a world build on IT and IT-security relying on cryptographic hash functions, it is easy to imagine the vast amount of things going wrong; from changing the letter of the law and sign it with a rouge European Commission certificate, change ownership of property to masquerading malicious code as an Apple software update or intercept and change an encrypted E-mail.

These have been outlined as plausible during this report, the purpose of this chapter is to quantify the damage this would have on society as a whole, imagining the three situations:

- General SHA-1 collisions (matching random data with random data) are doable within an hour
- 2nd pre-image SHA-1 collisions (matching specific data with random data) are doable within an hour
- 2nd pre-image SHA-1 collisions (matching specific data with specific data) are doable within an hour

There are several methodologies and standards that can be followed, with **NIST**, octave allegro being predominant ones in combination with **ALARP** and **SIL** from **IEC** 61508, where EN 50159²⁸⁵ sets up a more descriptive framework, close to that of **ALARP**.

The estimates in this report have been made with a mixture of those methods, not picking one specific approach and follow it 100%, but follow good suggestions and ideas coming from all of them.

13.1 Random Data Collision Within One Hour

-LARS

When **SHA-1** alone is used to insure integrity, such as is the case with BitTorrent and OpenPGP, being able to find a 2nd pre-image collision (with no objections against the input data being random), messages can be intercepted and the original datablock be removed and replaced with the collision.

With such a replacement, the message will still look authorised to the receiver, leading the contents to be changed without the receiver's knowledge.

The largest consequence for this is in the BitTorrent protocol, where a mangled piece (just 1000th of the download) can corrupt the whole download(in the case of compressed archives), without the protocol being able to detect it.

As more and more services use BitTorrent for content distribution networks and software update, this allows for a new kind of denial of service attack.

Denial of Service as the term is used by media today relies on filling bandwidth.

²⁸⁵ "Railway Applications - Communication, Signalling and Processing Systems - Safety-Related Communication in Transmission Systems - EN 50159."



Hash Denial of Service (HDOS) simply relies on replacing 1000th of the data with garbage that has the same digest as any part of the original data.

Worst case scenario:

Security updates to an operating system is corrupted leading to a known vulnerability being exploitable for a longer time. Corrupted files might also make embedded hardware crash as if random data passes the integrity check and is subsequently interpreted as executable code.

13.2 Specific Data Collision Within One Hour

-LARS

Exploiting the Merkle-Damgård structure of **SHA-1** the Initialisation Vectors can be pre-computed, not needing to process the whole document or code that can be several megabytes at each attempt given that there is a place in the end of the document that allows for 64 bytes of metadata padding, which is invisible to the user and is ignored by automated systems.

This allows subtle changes to code such as adding “or PWD=HardCodedBackdoor”, in text changing “with” to “without” or replace the RSA key in a certificate as illustrated in chapter 4 X.509 Structure, page 32. Complete rewrites are possible too, but subtle changes are harder to spot and can have huge consequences.

Worst case scenario:

- Firmware, software and updates certified by vendors can be modified to contain malicious software
- Law text cannot be verified, as long as **T**rusted **S**ervice **L**ists rely on the compromised hashing algorithm
- Digital signatures cannot be processed
- Previous contracts and agreements have to be re-verified as integrity is not assured
- Logs of work hours and access logs cannot be trusted without a non-repudiation guarantee

With the worst case scenarios in place, a risk evaluation can be performed.

14 Risk Evaluation

-LARS

With the hazards identified, the probabilities estimated, the impacts & consequences analysed it is now possible to assess the risk.

Previous chapters have shown that **SHA-1** 2nd pre-image attacks are high-consequence attacks, meaning that it should also be constrained to be a low probability scenario.

In this report it has been shown that 160 secure bits is outside of the scope of human wealth, meaning it is a possible goal to strive towards having authentication messages with a strength of 160 bits.

Exploring if the NIST recommendations for secure bits of 80(until 2014), 112(until 2031) and 128(2031+) would be nice, but was skipped due to time constraints.

Moreover, the estimates this is based on is a piece of software performing a factor of 10 worse than a test done in 2011, when normed on the number of nodes.



The graph below shows a comparison between Schneier, Stevens and Alexander:

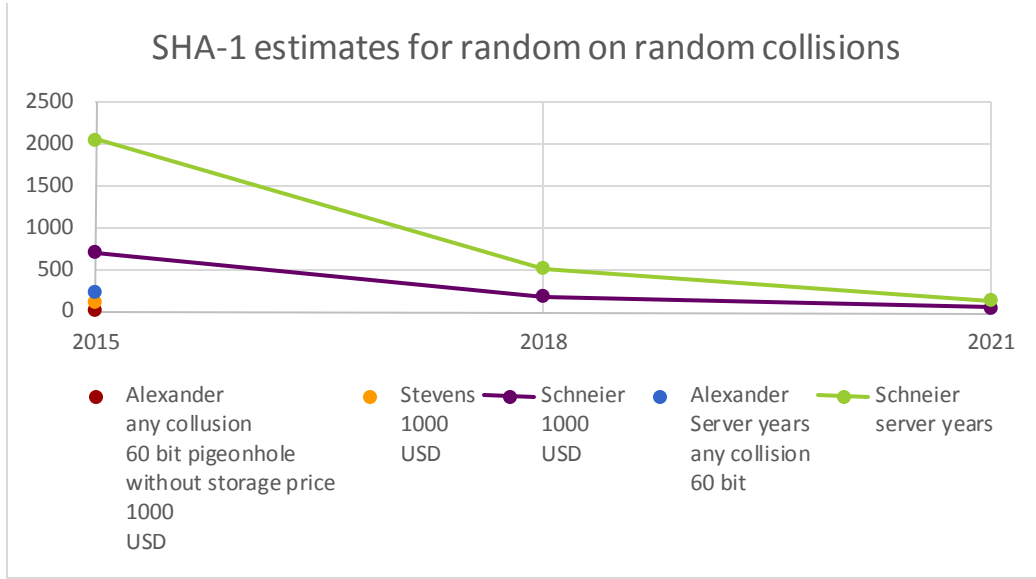


Figure 46 SHA-1 estimates for general collisions

The numbers from chapter 10.3 HPC Evaluation, page 63 showing the performance of the code is not shown, as on the scale of the graph above, those data points would be 1 to 23 billion light years away. The price figure by Alexander is without the price of storage, would Amazon EBS storage price²⁸⁶ be included, the data point should be featured 65 meters away.

Table 5 Comparison between estimates from Alexander, Schneier and Stevens

Year	Alexander any collision 60 bit birthday without storage price 1000 USD	Stevens 1000 USD	Schneier 1000 USD	Alexander Server years any collision on 60 bit	Schneier server years	Alexander 2nd pre-image collision on 1000 USD	Alexander any collision on 1000 USD	Alexander any collision 80 bit birthday with storage price 1000 USD	Alexander any collision 60 bit birthday with storage price 1000 USD
2012			2700		8192				
2015	17	98	700	218	2048	1,488 E+31	9,313 E+29	4,352 E+12	4,415 E+6
2018			173		512				
2021			43		128				
Distance to data point in km (on the scale of Figure 46)						2,17 E+23	1,36 E+22	63750	0,065
billion light years						23	1,4		
% of the diameter of the observable universe						25	1,6		
times the distance to the most distant quasar (SDSS J1148+5251 at a redshift of 6.41)						1,8	0,11		

²⁸⁶ "AWS | Amazon EBS | Pricing."



14.1 Schneier misunderstanding Stevens

-LARS

In the article “When Will We See Collisions for SHA-1?”²⁸⁷ Schneier emphasis:

“practical collision attack against SHA-1”

“A collision attack is therefore well within the range of what an organized crime syndicate can practically budget by 2018”

“The point is that we in the community need to start the migration away from SHA-1 and to SHA-2/SHA-3 now.”

Quoting Marc Stevens²⁸⁸ as the source for probabilities.

Stevens’s article deals with general “random on random” general collisions, that have little impact and only academics, but no crime syndicate would be interested in.

As Stevens write:

“Collisions on SHA-1 can result in signature forgeries, but do not directly undermine the security of the Internet at large. More advanced so-called chosen-prefix collisions are significantly more threatening, but currently much costlier to mount. Yet, given the lessons learned with the MD5 full break, it is not advisable to wait until these become practically possible.”

-Marc Stevens,²⁸⁹

From Schneier’s article it seems like this has been misunderstood to be a 2nd pre-image attack (specific on specific) **SHA-1** attack, one that crime syndicates indeed would be very interested in forging due to many opportunities for profit as described in earlier chapters.

This also explains the discrepancy in data, with Schneier’s figures describing the much less computational intensive “random on random” general collision and not the 2nd pre-image “specific on specific” collision.

14.2 Analysis on the estimates derived from own data

-LARS

The reliance on 160 secure bits is sound.

If all 160 bits of **SHA-1** were secure the only way was to brute-force a 2nd pre-image attack, getting a 50% chance of success would take $1,488 \times 10^{34}$ **USD** in electricity alone, and a general collision 9.313×10^{32} **USD** (2015 price, calculations by Alexander Brandbyge, derived from energy consumption of **HPC** running the code described in chapter 7 GPU SHA-1 Collision Probability Estimate, pages 49-60).

For a comparison all the **USD** in circulation in the world amounts to **USD** 1,39 trillion²⁹⁰; $1,39 \times 10^{12}$ **USD** meaning that a 50% chance of a 2nd pre-image attack on 160 secure bits would take 10’000’000’000’000’000’000’000 times more than the amount of **USD** in the world.

Hence it still seems infeasible to produce a 2nd pre-image collision, but if the goal just is to find a general **SHA-1** collision of 60²⁹¹ bits, due to the pigeonhole principle the price for **CPU** time will only be 17’030 **USD** in power, or 1’243’190 **USD** in rent on Amazon EC2, but would furthermore require at least $4,15 \times 10^7$ TeraBytes (calculations by Alexander Brandbyge)

²⁸⁷ Schneier, “When Will We See Collisions for SHA-1? - Schneier on Security.”

²⁸⁸ Marc, “Cryptanalysis of MD5 & SHA-1.”

²⁸⁹ Stevens, Karpman, and Peyrin, “Freestart Collision on Full SHA-1.”

²⁹⁰ “FRB: How Much U.S. Currency Is in Circulation?”

²⁹¹ Stevens, “Cryptanalysis of MD5 & SHA-1.”



Costing 0,1 USD for each GB²⁹², it amounts to 4,15 billion USD, in storage rent, not accounting for the fact that it is over 20'000 times more storage than the capacity of the supercomputer ranked number 267 in the world²⁹³.

Having shown that readily available 2nd pre-image collisions is a high consequence scenario, it is comforting that Alexander's results shows it to be a classic catastrophic consequence, low probability event.

Cation should be taken though, as with metal fatigue the probability increases each year, not due to wear and tear, but Moore's law doubling the computing power available for 1USD each 2½ years²⁹⁴ and the steady discovery of vulnerabilities to SHA-1.

15 Risk mitigation: Responsible Disclosure

-LARS

As mentioned in chapter 2.2.1.2 Responsible Disclosure in a Risk Assessment Perspective, page 18, near miss and bug reporting is theoretically an effective tool.

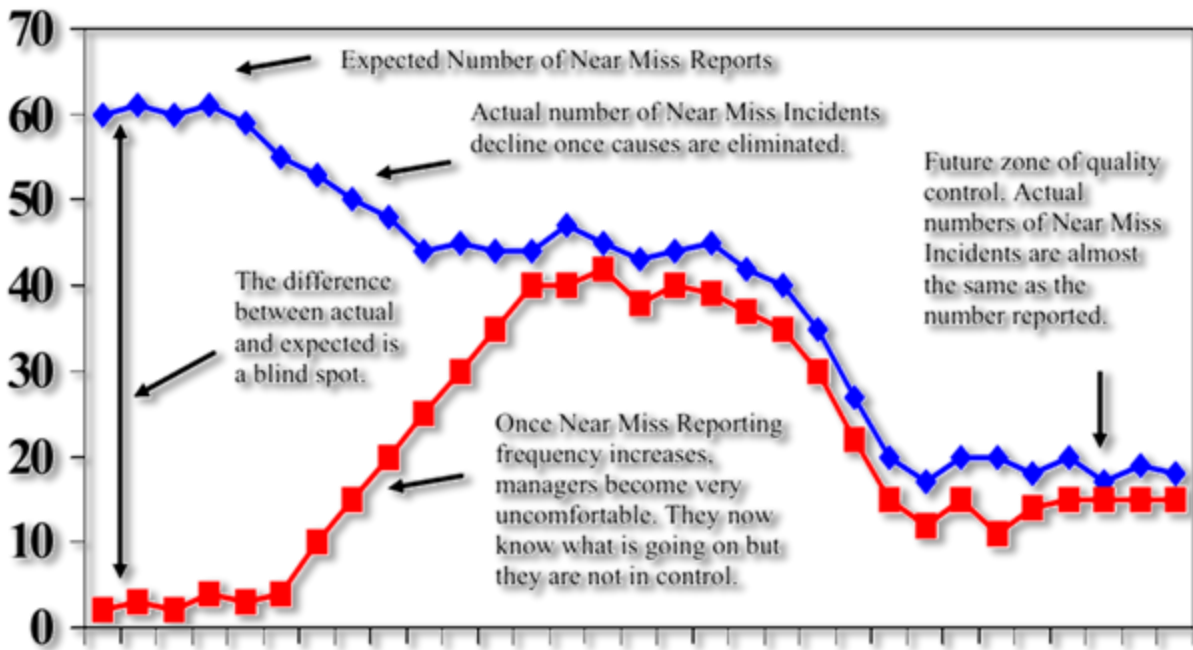


Figure 47 Expected and actual Near Miss ratio, by ²⁹⁵

Figure 47 above shows how welcoming incident reporting will lead to large increase in reports, making it easy for management to panic during such a campaign, especially for companies traded at the stock market, as the number of reports will increase dramatically and yearly statistical reports will make it seem like the company is performing worse than earlier.

A lenient approach will have to deal with a lot of reports already covered by company rules as being reasons for termination.

But in the long run a lenient approach will lead to fewer incidents as seen in Figure 49 below.

²⁹² "AWS | Amazon EBS | Pricing."

²⁹³ "TOP500 Supercomputer Sites | 267."

²⁹⁴ Clark, "Intel Rechisels the Tablet on Moore's Law."

²⁹⁵ Borg, "Predictive Safety from Near Miss Hazard-Reporting."

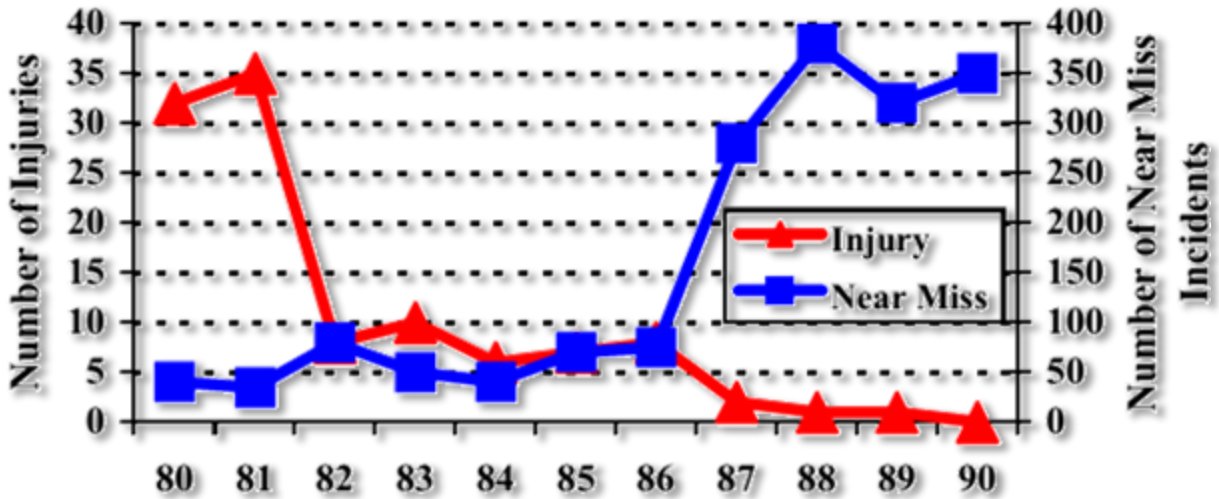


Figure 48 Effect of Near Miss reporting on injuries at a major petroleum company in Canada in the 1980s, by ²⁹⁶

An effect of this trend can be seen in the computer security domain during the aftermath of *Heartbleed*; that while being published as an issue on the 7th of April and receiving much publicity, researchers still found vulnerable servers in the end of April.²⁹⁷

The 28th of April researchers sent notification messages to some of the server owners to let them know they were vulnerable and sent another batch of messages the 7th of May.

The number of servers patched is shown below to illustrate the significant difference in those who have received a notification and those who got it a week later.

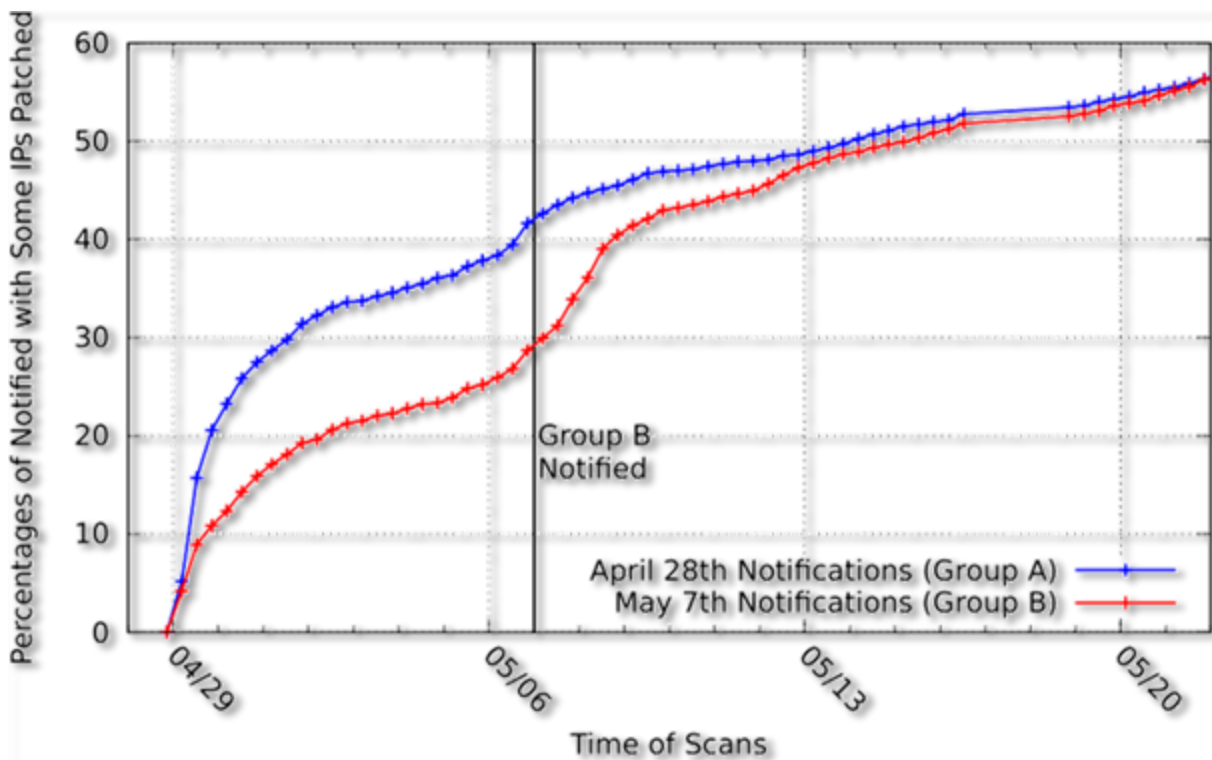


Figure 49 Difference between notified and un-notified servers, by ²⁹⁸

This should be seen in contrast to the generally fast response to major publicised security vulnerabilities seen below.

²⁹⁶ Ibid.

²⁹⁷ Durumeric et al., "The Matter of Heartbleed."

²⁹⁸ Ibid.

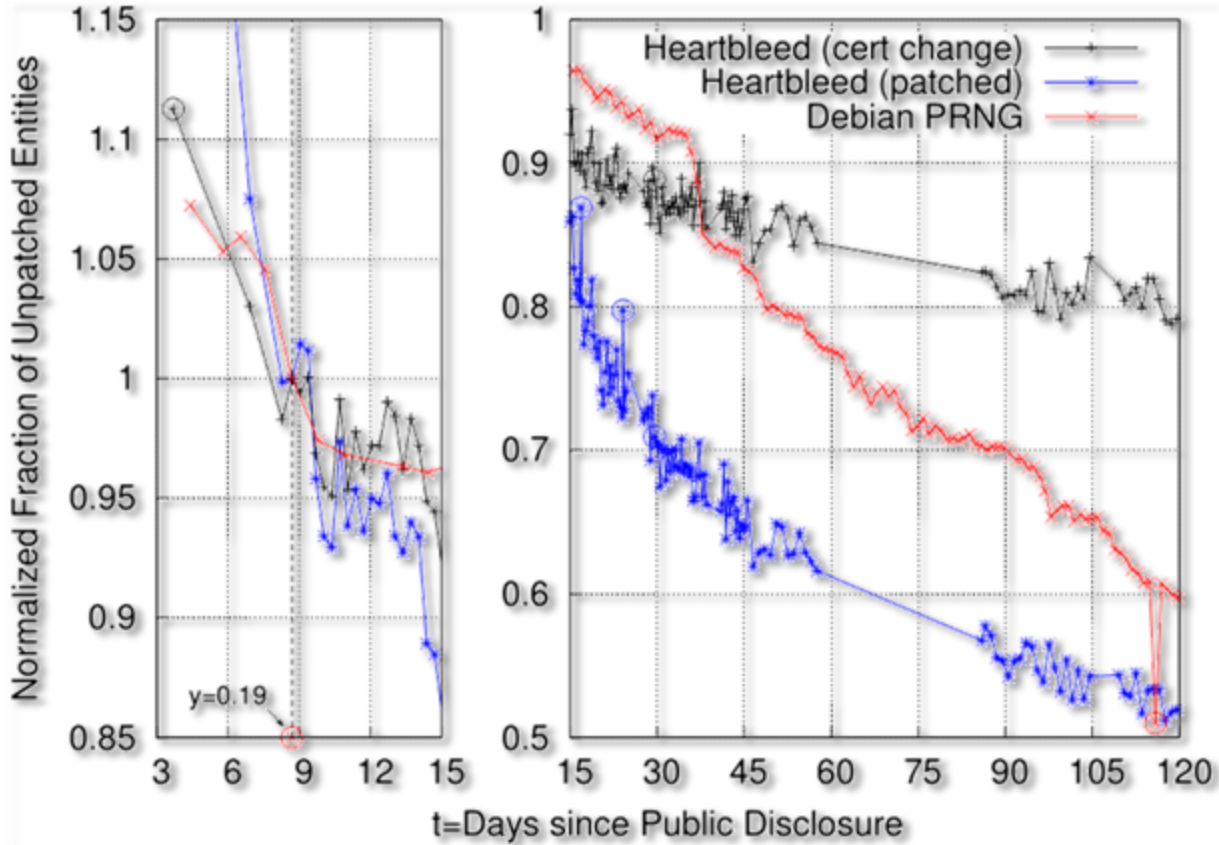


Figure 50 Historical response to security incidents, by ²⁹⁹
 In the domain of software security and bug reporting there have been a documented tendency to either ignore or incriminate those providing reports³⁰⁰.

An example of incriminating security research is a recent Danish court case where a person was convicted for “accomplice in attempted hacking” (getting convicted to 6 months of jail, after having spent 16 months in pre-trial jail), leading to the precedence that talking about security issues can in itself be illegal, if the person you talk with will then test the theory, even if they fail to penetrate or break any system (attempted hacking). ^{301, 302}

In the private sector here is a new trend of rewarding user submitted reports on vulnerabilities among larger international companies the so called bug bounties where companies pay in cash goods or services for detailing security bugs, neglects and overall attack vectors able to penetrate live services. ³⁰³

²⁹⁹ Ibid.

³⁰⁰ The European parliament and council, *Directive 2013/40/EU (Cybercrime)*.

³⁰¹ Conviction in the case of hacking of CSC (municipal court of Frederiksberg 2014).

³⁰² Transcript of hacker case (municipal court of Frederiksberg 2014).

³⁰³ “The History of Bug Bounty Programs.”

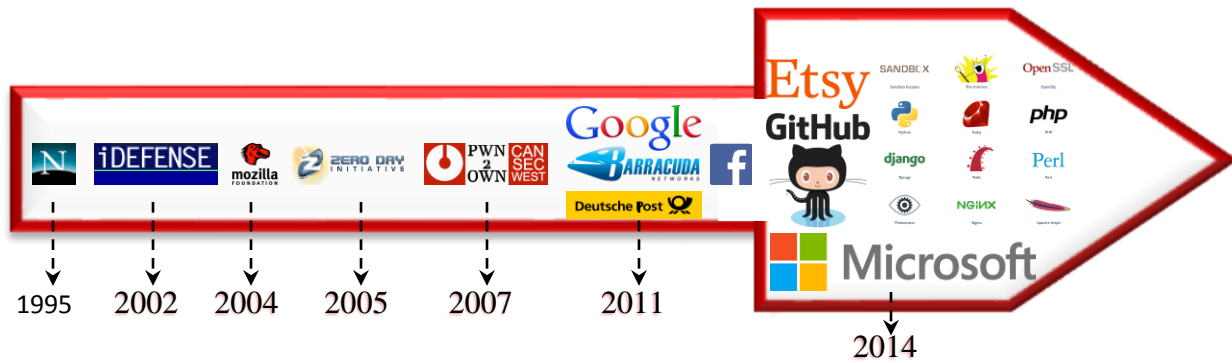


Figure 51 History of Bug Bounty programs, by Lars Embøll

Netscape launched the first bug bounty program in 1995³⁰⁴, but was first in the 2000s that more companies adapted it and by 2014 became the standard way to respond to *Near Misses* in the form of reports of vulnerabilities.^{305, 306, 307}

Since 2004 bug bounty platforms have emerged, disrupting the way vulnerability reports are handled; they relieve the security researcher from having to contact software vendors, set up secure connections for data and manage payment of the bounties. While most bug bounty platforms merely act as a secretary doing administrative functions, some goes further doing background checks and verifying the vulnerabilities.

There is no standard on the time between submission and public disclosure, it varies from 45 days for *CERT*, to 120 days for the *Zero Day Initiative*, with the 2014 google program *Project Zero* having a 90 days hard deadline.^{308, 309, 310}

The *HackerOne* bug bounty platform has a more lenient approach and publication varies from the default 30 days to 180 days for uncooperative vendors; with 30 days being the goal, but accepting vendor timelines up and until 180 days.

While it is discouraged, the legal terms of one of the biggest bug bounty platforms; *HackerOne* enables security researchers to sell the vulnerabilities to multiple bug bounty platforms, the dark web and to disclose it to the public, as there is a non-exclusivity rule.

"You grant HackerOne a non-exclusive, worldwide, perpetual, irrevocable, royalty-free, fully paid-up, sublicensable and transferable right to use, copy, reproduce, display, modify, adapt, transmit, and distribute the Content, in any media now known or not currently known, for any business purpose."

A way to deter premature disclosure is to confer ownership of the exploit to the bug bounty platform owners, an example is the *Zero Day Initiative* that specify:

*"Any code execution vulnerability that the Zero Day Initiative awards a cash prize for becomes the property of the ZDI, and therefore the winner cannot discuss or disclose details of the 0-day until the affected vendor has successfully patched the issue."*³¹¹

This benefits the owner of the *Zero Day Initiative*; HP DV Labs (Hewlett-Packard Digital Vaccine Labs) as:

³⁰⁴ "Netscape Bugs Bounty."

³⁰⁵ "Microsoft and Facebook Launch Internet Bug Bounty Program."

³⁰⁶ "HackerOne: Vulnerability Coordination and Bug Bounty Platform."

³⁰⁷ "The History of Bug Bounty Programs."

³⁰⁸ "Vulnerability Disclosure Policy | Vulnerability Analysis | The CERT Division."

³⁰⁹ "Zero Day Initiative - Disclosure Policy."

³¹⁰ "Project Zero: Announcing Project Zero."

³¹¹ "Zero Day Initiative - Disclosure Policy."



“DVLabs may distribute vulnerability protection filters to its customers' IPS devices through the Digital Vaccine service”.

Utilizing the ownership to sell protection from otherwise unknown vulnerabilities.

The income for the HackerOne platform is 20% of the bounties paid out to the researchers, relying on the economy of scale in centralizing contact and administration, rather than having each individual security researcher contacting the vendors.³¹²

15.1 Storing Secrets Securely

-LARS

In an effort to comply with the European cybercrime directive Article 8³¹³, potentially harmful data must be kept in a state where it is unable to interfere or disrupt telecommunication infrastructure. If the data would in any way incite or aid an attempt of *“seriously hindering or interrupting the functioning of an information system by ... altering ... data”*³¹⁴ (Article 4) it would be a criminal act.

15.1.1 Shamir Secret Sharing

-LARS

Shamir secret sharing is a way of splitting a secret into multiple pieces (shards) so that more than one piece(shard) is needed to be combined in order to extract the secret³¹⁵. This is done by transforming the data into multidimensional planes that intersect at specific points and only with multiple of these fields available the right intersection points and the original data can be found.

The drawback is that the shards have almost the same size as the secret, but on their own they are just random data.

In order to comply with the law and limit the probability of misuse the **RSA** key for the forged European Commission certificate (that can be seen in appendix 20.1.1, page 115) was split into 5 shards, with the need for at least 3 to be combined to extract the key again.

Meaning that if there had been a 2nd pre-image collision no single person could use the key to sign documents on behalf of the European Commission with the forged certificate, it would need 3 people to be present.

In the case of illness or death of an author 2 shard holders could be summoned and would be able to re-create the key with the remaining author if needed, with the shards acting as safe backups distributed throughout the Nordic countries.

15.1.2 Setup

-LARS

The chosen field and software is $GF(2^8)$ ³¹⁶ and libgfshare³¹⁷.

The software was run in a Virtual Machine on a freshly formatted air-gapped computer running ESXi 6.

The client computers were also freshly formatted and air-gapped, running Kali Linux getting the dependencies from USB pen or DVD.

- 1) The setup script was written, tested and run in this environment and did the following:
Install the dependencies from local storage (with the github links available, but outcommented)
- 2) Generate the certificate key and save it in a file

³¹² “Terms of Service - HackerOne.”

³¹³ The European parliament and council, *Directive 2013/40/EU (Cybercrime)*.

³¹⁴ Ibid.

³¹⁵ McVittie, “Theory Used by Libgfshare.”

³¹⁶ Ibid.

³¹⁷ “Djpholy/libgfshare.”



- 3) Sign the original European Commission certificate with the key to verify existence and ownership of private key by signing a "Nothing up my sleeve"³¹⁸ value
- 4) Split the key into 5 shards
- 5) Test that permutations of 3 shards can recreate the key, but that 2 or a single one cannot.
- 6) Securely delete the key (200 passes, ending with a 201st pass consisting of zeros)
- 7) Encrypt each shard with the public OpenPGP key of the precipitants
- 8) Securely delete each shard once encrypted (200 passes, ending with a 201st pass consisting of zeros)
- 9) Sign each shard to verify integrity and sender

Each shard was then loaded into individual USB pens (that were bought from a physical shop using cash and freshly formatted at one of the air-gapped computers)

Then immediately hand delivered to the recipients while making sure that they were under the supervision of at least 2 people while 3 or more keys were at the same place (even though the contents is encrypted).

15.1.3 Other usage

-LARS

It is a bit precautionary to use Shamir secret sharing in order to store the key of a forged certificate that has its complete ASN.1 code in several countries' **TSL**, on the other hand it is a practical exercise in good security. Instances where it could be of use is where a high value key is used sparingly, an example could be the major updates of an OS like OS X and iOS where a key could be generated to sign the key, then deleted or split ensuring that no one else could generate that signature, while relying on less critical keys for intermediate updates.

16 Summary of Part 3

-LARS

As it has been shown there are several instances of older hashing algorithms being used.

It has also been shown that searching for bugs and managing vulnerabilities has proven to be a complex problem for even the largest of companies on this planet.

A mitigation to this seems to be an effort to crowdsource vulnerability reports through in-house or 3rd party bug bounty systems.

While companies still have a legal responsibility to keep their services secure, bug bounty systems provide an opportunity to expand the knowledge of previously unknown vulnerabilities.

For government entities and pan-national standards an *open proof* approach can ease the understanding and third party testing of security.

Along with a culture welcoming incident reports from 3rd parties interested in security.

³¹⁸ "Sha 1 - Why Initialize SHA1 with Specific Buffer?"



PART 4 CONCLUSION

17 Conclusion

A **HPC** application was developed and tested, and while it was not capable of generating a valid forged certificate, it successfully provided a benchmark of the brute-force generation rates attainable by the ABACUS 2.0 **GPU** Nodes, and by extension what is possible with current hardware. An estimate of the efficiency of known **SHA-1** optimization techniques and cryptographic attacks has also been performed using this **HPC**.

With the overarching goal of updating the price estimates of Schneier³¹⁹ this report has produced a figure comparatively in **CPU** price with the Stevens³²⁰ estimate for a general collision with 60 secure bits, but also a new figure: The lower bound for the price of 50% chance for a 2nd pre-image (specific on specific), which is $1,488 \times 10^{34}$ **USD** or 10'000'000'000'000'000'000'000'000 times more than the amount of **USD** in the world.

Secondly it has been established that precautionary measures in the order of **SIL 4** should be taken regarding equipment using **SHA-1** (or low entropy authentication messages below **NIST SP 800-57**³²¹ recommendations) doing safety related tasks, as the consequences are catastrophic and though experts disagree on the specific timeframe, they all caution a change away from **SHA-1** as the first hints of a broken algorithm appears and better alternatives are tested and available in the form of **SHA-3**.

Thirdly, the method of forging an X.509 certificate has been reproduced and verified, as done by Stevens in 2009³²².

This was a blind test, as the method was devised and tested, before the article of Stevens was revealed, further strengthening it as a good target for 2nd pre-image attacks, illustrating the current reliability of hashing for security.

Finally, while the **BitTorrent** application did prove successful in gathering and analysing a significant amount of torrent metadata files, the results were clear; the amount of **SHA-1** data available in the entire **BitTorrent** network is simply not enough to be useful as a rainbow table and the **BitTorrent** network as such has no impact on the security of the **SHA-1** function. Furthermore, the **BitTorrent** protocol is not in any specific risk of collision attacks, since the piece sizes are spread across a large set of values and the amount of pieces in each piece size group is insignificant.

³¹⁹ Schneier, "When Will We See Collisions for SHA-1? - Schneier on Security."

³²⁰ Stevens, Karpman, and Peyrin, "Freestart Collision on Full SHA-1."

³²¹ Barker, "Recommendation for Key Management: Part 1: General (Revision 4) DRAFT SP800-57."

³²² Stevens et al., "Short Chosen-Prefix Collisions for MD5 and the Creation of a Rogue CA Certificate."



17.1 Recommendations

17.1.1 Future projects should use SHA-3

-LARS

With time **SHA-1** will be phased out and with Microsoft³²³, Google³²⁴ and the European research and education network TERENA/Géant³²⁵. Recommending a deprecation of **SHA-1** from December 31st 2015 it is unadvisable to incorporate **SHA-1** into future projects, especially ones with a long life-cycle.

As it is mentioned in Octave Allegro:

*The profiling process establishes clear boundaries for the asset, identifies its security requirements, and identifies all of the locations where the asset is stored, transported, or processed.*³²⁶ p. 17

Emphasis has been added to the second part, illustrating an information management approach scaling security level and cost to the asset in order to avoid high unnecessary marginal costs associated with applying high risk mitigation to a broader system.

While it is possible to outline all systems using or depending on **SHA-1** in the event of 2nd pre-image collisions being possible in 10 years by a quantum leap from a disruptive technology the additional investment for **SHA-3** compared to **SHA-1** is neglectable, only using computer resources (cycles) under a factor of 3 more than **SHA-1** while providing 256 secure bits compared to < 80 for **SHA-1**³²⁷.

(12,6 cycles/byte³²⁸ (PAGE 25) for **SHA-3**, versus 4,32 cycles/byte³²⁹ for **SHA-1**)

As the security doubles for each additional secure bit the increased security gotten by this is astronomical.

In order to be worth the extra resources needed, **SHA-3** would only have to feature 2 more secure bits, but while the exact increase of secure bits is unknown it is at least 176 more, resulting in a security 2^{176} or 9.5×10^{52} times better.

17.1.2 Authentication Message Entropy

-LARS

The authors also suggest a continuous increase of the number of secure bits; rather than increasing it in steps of 32 or 16 bits (80-112 in 2014, 112-128 in 2030)³³⁰, there could be an increase in the number of required secure bits each year.

Since increasing with a bit means doubling the outcome space, it will combat Moore's law that "only" doubles each 2 to 2½ years.

³²³ "SHA1 Deprecation Policy - Windows PKI Blog - Site Home - TechNet Blogs."

³²⁴ "Intent to Deprecate: SHA-1 Certificates - Google Groups."

³²⁵ "TERENA> News> TCS CertificateService Responds to SHA Security Update."

³²⁶ Caralli et al., "The OCTAVE Allegro Guidebook, v1. 0."

³²⁷ Stevens, Karpman, and Peyrin, "Freestart Collision on Full SHA-1."

³²⁸ Guido et al., "Keccak Implementation Overview."

³²⁹ "Measurements of Hash Functions, Indexed by Machine."

³³⁰ Barker et al., "Recommendation for Key Management SP 800-57 Part 1: General Revision 3," 57.



Giving the following progression, making it easy to implement, plan and maintain long term systems:

Year	Required number of secure bits	Disallowed
2030	130	116
2029	129	115
2028	128	114
...	Years after 2000 + 100	Years after 2000 + 100-14
2020	120	106
2018	118	104
2016	116	102
2014	114	100

Suggestion for projection of number of secure bit requirements submitted to **NIST** as comment for SP 800-57 draft.

Following this suggestion, even if it is not amended to the **NIST** SP 800-57 is a good rule of thumb that will keep projects within (and a bit above) current **NIST** recommendation.

17.1.3 OpenPGP RFC 4880

-LARS

It is our recommendation that it should be updated to not mark SHA-1 as mandatory in **SECTION 9.4.** and look into the current use and possible exploitation of the field specified in **SECTION 13.3.2.** that allows the sender to specify the hashing algorithm they want the recipient to use for replies.

17.1.4 Certificate Transparency

-LARS

Certificate transparency is a good way to move away from the reliance of single hashes to provide the authentication of certificates and is already implemented in Google (Alphabet) projects³³¹.

CatLfish³³², the list of known Certificate Transparency logs³³³ and the guide on how to manually verify Signed Certificate Timestamp with openssl³³⁴ are good places to start.

17.1.5 Flexibility in security critical container types

-ALEXANDER

An important property of the Merkle-Damgård construction is that it allows for the construction of pre-hashes. It works by breaking down the variable sized input, to fixed size chunks and then applying the compression function to each chunk, before finally combining the chunks in order.

By allowing variable sized input in security critical types, such as allowing a comment field in the X.509 certificate structure to exist, allows for a potentially severe reduction in work in brute force cases.

When attempting to perform a brute force collision on this container, all preceding/following chunks can be calculated in advance result can be reused, thereby only needing to do a single compression per brute-force trial instead of multiple.

³³¹ "Certificate Transparency in Chrome - Certificate Transparency."

³³² "Certificate Transparency Playground."

³³³ "Known Logs - Certificate Transparency."

³³⁴ "Certificate Transparency."



Therefore, for containers such as the X.509 certificate, variable sized inputs should be disallowed or if possible, not contribute to the main signature. If that is impossible, they should at the very least be forced to reside in such a way that they cannot possibly fit within the bounds of a single chunk.

Alternatively, Hash functions built on the Merkle-Damgård construction, or a similar architecture, should not be used for this type of data.

17.1.6 Tip on Good Hash

-LARS

The **NIST** competition for **SHA-3** chose Keccak as the winner, but the competition still had 3 years of thorough investigation on a lot of promising candidates, several of them with no found weaknesses.

While security by obscurity is to use undocumented algorithms in the hope that it will make the work harder for an intruder, choosing one of the **NIST** finalist³³⁵ is the exact opposite; heavily documented and tested cryptographic hash functions but with a much smaller attack surface than Keccak due to a smaller user-base.

http://csrc.nist.gov/groups/ST/hash/sha-3/Round3/submissions_rnd3.html

Thanks goes out to the researchers behind these candidate algorithms.

17.2 Future Work

17.2.1 HPC

-ALEXANDER

While the **HPC** application performed the job of brute-forcing the X.509 Certificate, the project left several avenues of improvements available.

Primarily, having the application perform two jobs at once, both brute-forcing the certificate as well as testing between warps for a general collision turned out to be superfluous. Splitting the forcer in two distinct applications with each their purpose would serve as a better use of computation time as improve the scalability of the code.

Secondly, while serving as a good **SHA-1** reference implementation, much more work could be done in optimizing the code to better exploit **CUDA** intrinsic instructions as well as just general high performance code optimization. This could easily take up an entire new thesis however.

Also, the different attacks made against the algorithm could be exploited to drastically reduce the computations needed to generate a general collision, so these would a prime target for further elaboration.

In relation to the optimizations, a more thorough diagnostics of the performance would be welcome, since the only metric used were the count of digests generated per second. The **CUDA** toolkits offer powerful application diagnostics which allows a developer to monitor every level of memory consumption as well as estimating the occupancy and how to improve the application. It could be useful to spend some time with this, as there might be untapped resources still on the GK110 chips.

17.2.2 Torrent

-ALEXANDER

The Torrent Application ended up primarily working on a large bank of magnet links fetched out-of-band. This means it lacks the capabilities to truly communicating with a torrent swarm, and in order to improve the accuracy of the measurements this would be the logical next step.

³³⁵ "THIRD (FINAL) ROUND SHA-3 CANDIDATES - NIST.gov - Computer Security Division - Computer Security Resource Center."



Also by doing this, the application would gain the ability to react to changes in the input dataset and perhaps even be used as a staging ground for torrent based attacks.

Lastly, the fact that an abnormal amount of torrents with the exact piece count of 1088, remains a complete mystery. Why this occurs, even though there is nothing in the standard to suggest this should be or any documentation on the subject, is simply unknown and as such could be an interesting subject to explore.

17.2.3 Data on SHA-1 usage

-LARS

Suggested future would be to analyse the current use of **SHA-1** in **SSL/TLS** certificates, preferably with a comparison for data the months before December 31 2015 and **Heartbleed**. Looking into the Alexa top 1000, top 1 million compared to personal websites using self-signed certificates and the new free certificate services. A viable method could be to use the certificate transparency protocol specified in **RFC 6962** and the list of publicly available servers at <http://www.certificate-transparency.org/known-logs> as well as <https://plausible.ct.nordu.net>, compared to <https://www.trustworthyinternet.org/ssl-pulse/>, eg using:

```
curl -o certlog.log "https://<log server>/ct/v1/get-entries?start=0&end=X"

echo -n | openssl s_client -connect HOST:PORTNUMBER | sed -ne '/-BEGIN CERTIFICATE-/,/-END CERTIFICATE-/p' > /tmp/$SERVERNAME.cert
```

There is also the possibility of looking into the PGP strong set and the OpenPGP RFC 4880 and Google's End to End:

*I've never looked at the "Hash Algorithm" stuff,
and I don't see how to (when given a KEY BLOCK),
get the info with simple gpg-commands.*

*The wotsap stuff doesn't store the key-blocks ;
it fetches a block, looks at the sigs, and then
removes the block.*

*To get a list of key-ids, you can use my stuff ...
<http://www.staff.science.uu.nl/~penni101/wotsap/>
... to decompress, unpack etc wotsap-archives from
<http://pgp.cs.uu.nl/archive/>*

*Regards,
Henk Penning*

- Henk Penning, author of "analysis of the strong set in the PGP web of trust", <http://pgp.cs.uu.nl/plot/>

As **RFC4880 SECTION 13.3.2.** details, that for practical reasons a sender can specify the hashing algorithm they want the recipient to use for replies e.g. an older weaker hashing algorithm.

This opens up for a downgrade attackvector weakening the security to at least **SHA-1**.

With **SHA-1** being the mandatory default, **SHA-1** is currently the fall-back if nothing is specified leading to most software not specifying a hashing algorithm.

"Since SHA1 is the MUST-implement hash algorithm, if it is not explicitly in the list, it is tacitly at the end. However, it is good form to place it there explicitly."

-RFC4880 SECTION 13.3.2.

We recommended a revision of the **RFC** and implementation to include the request for stronger hashing algorithms.



18 Bibliography

- 3rd, D. Eastlake, and Paul E. Jones. "RFC3174 - US Secure Hash Algorithm 1 (SHA1)." Accessed July 20, 2015. <https://tools.ietf.org/html/rfc3174>.
- "286.pdf." Accessed December 20, 2015. <http://eprint.iacr.org/2011/286.pdf>.
- "Abacus 2.0 | DelC National HPC Centre, SDU." Accessed December 26, 2015. <https://deic.sdu.dk/>.
- Abadie, Andre', Damindra Bandara, and Duminda Wijesekera. "A Composite Risk Model for Railroad Operations Utilizing Positive Train Control (PTC)," V001T06A004. ASME, 2014. doi:10.1115/JRC2014-3730.
- "About Us | Bitsnoop." Accessed December 23, 2015. <http://bitsnoop.com/info/about.html>.
- Adams, John. "The Economics and Morality of Safety Revisited," 2009. <http://john-adams.co.uk/wp-content/uploads/2009/02/teamos.pdf>.
- A Family Tree for Humanity*. Accessed December 21, 2015. http://www.ted.com/talks/spencer_wells_is_building_a_family_tree_for_all_humanity?language=en.
- "Akamai: Gamers Aren't P2P Bandwidth Slaves - TorrentFreak." Accessed January 2, 2016. <https://torrentfreak.com/akamai-gamers-arent-p2p-bandwidth-slaves-100915/>.
- A. M. de Bruin, René Bekker. "Dimensioning Hospital Wards Using the Erlang Loss Model. *Ann Oper Res.*" *Annals OR* 178, no. 1 (2010): 23–43. doi:10.1007/s10479-009-0647-8.
- "Analysis of the Strong Set in the PGP Web of Trust." Accessed July 28, 2015. <http://pgp.cs.uu.nl/plot/>.
- Andrews, Rick. "The Cost of Creating Collisions Using SHA-1." *CA Security Council*. Accessed June 30, 2015. <https://casecurity.org/2014/11/18/the-cost-of-creating-collisions-using-sha-1/>.
- "Anonymous Hacker Group: Two Jailed for Cyber Attacks." *BBC News*. Accessed December 26, 2015. <http://www.bbc.com/news/uk-21187632>.
- "Anonymous Hackers 'Cost PayPal £3.5m.'" *BBC News*. Accessed December 26, 2015. <http://www.bbc.com/news/uk-20449474>.
- Apollo Reliability and Quality Assurance Office. "Procedure for Failure Mode, Effects and Criticality Analysis (FMECA)." National Aeronautics and Space Administration, August 1966. http://www.fmeainfocentre.com/handbooks/19700076494_1970076494.pdf.
- Argyros, George, and Aggelos Kiayias. "PRNG: Pwning Random Number Generators," 2012. https://media.blackhat.com/bh-us-12/Briefings/Argyros/BH_US_12_Argyros_PRNG_WP.pdf.
- "Aviation Safety Network > ASN Aviation Safety Database > Aircraft Type Index." Accessed January 3, 2016. <http://aviation-safety.net/database/type/index.php>.
- "AWS | Amazon EBS | Pricing." *Amazon Web Services, Inc.* Accessed January 9, 2016. <http://aws.amazon.com/ebs/pricing/>.
- Barker, Elaine. "Recommendation for Key Management: Part 1: General (Revision 4) DRAFT SP800-57." National Institute of Standards and Technology. Accessed September 25, 2015. http://csrc.nist.gov/publications/drafts/800-57/sp800-57p1r4_draft.pdf.
- Barker, Elaine, William Barker, William Burr, William Polk, and Miles Smid. "Recommendation for Key Management SP 800-57 Part 1: General Revision 3." *NIST Special Publication* 800, no. 57 (2007): 1–142.
- Barker, Elaine B., and Allen L. Roginsky. "Transitions: Recommendation for Transitioning the Use of Cryptographic Algorithms and Key Lengths SP 800-131 A Rev. 1." National Institute of Standards and Technology, November 2015. <http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-131Ar1.pdf>.
- Bellare, Mihir, and Sara K. Miner. "A Forward-Secure Digital Signature Scheme," 431–48. Springer-Verlag, 1999.
- "BitTorrent Goes Legit with Content Delivery Service - InternetNews." Accessed January 2, 2016. <http://www.internetnews.com/xSP/article.php/3704076>.
- "Bombardier Enters ERTMS Level 2 High Speed Rail Control Market in Spain - Bombardier." Accessed December 5, 2015. <http://www.bombardier.com/en/media/newsList/details.bombardier-transportation20140613bombardierentersertmslevel2high.bombardiercom.html>.
- Borg, Bernard. "Predictive Safety from Near Miss Hazard-Reporting," 2002. <http://signalsafety.ca/files/Predictive-Safety-Near-Miss-Hazard-Reporting.pdf>.
- Caralli, Richard A., James F. Stevens, Lisa R. Young, and William R. Wilson. "The OCTAVE Allegro Guidebook, v1.0." *Software Engineering Institute*, 2007.
- "Certificate Transparency in Chrome - Certificate Transparency." Accessed January 11, 2016. <https://www.certificate-transparency.org/certificate-transparency-in-chrome>.
- "Certificate Transparency: Manually Verify SCT with Openssl." *Pierky's Blog*. Accessed October 23, 2015. <http://blog.pierky.com/certificate-transparency-manually-verify-sct-with-openssl/>.



- “Certificate Transparency Playground.” Accessed October 26, 2015. <https://www.ct.nordu.net/>.
- Christensen, Clayton M. *The Innovator’s Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business School Press, 1997.
- Clark, Don. “Intel Rechisels the Tablet on Moore’s Law.” *WSJ Blogs - Digits*, July 16, 2015. <http://blogs.wsj.com/digits/2015/07/16/intel-rechisels-the-tablet-on-moores-law/>.
- Cohen, Bram. “The BitTorrent Protocol Specification.” Html. *The BitTorrent Protocol Specification*, October 11, 2013. http://www.bittorrent.org/beps/bep_0003.html.
- Collins, Robert L. “Heinrich’s Fourth Dimension.” *Open Journal of Safety Science and Technology* 01, no. 01 (2011): 19–29. doi:10.4236/ojsst.2011.11003.
- “Combinatorics (2.6) The Birthday Problem (2.7) - bday_14-Handout.pdf.” Accessed September 8, 2015. http://www.math.ucsd.edu/~gptesler/186/slides/bday_14-handout.pdf.
- Commission Directive 2009/149/EC Common Safety Indicators (Definitions of an Accident). Vol. Commission Directive 2009/149/EC. Accessed April 16, 2013. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:313:0065:0074:EN:PDF>.
- Commission Directive 2009/149/EC (Definitions of an Accident). Vol. Commission Directive 2009/149/EC. Accessed April 16, 2013. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:313:0065:0074:EN:PDF>.
- “Common Safety Indicators, Denmark 2010, Version 5, Validated.” Accessed December 27, 2015. <https://erail.era.europa.eu/csi-data.aspx?country=5&year=2010&public=1>.
- “Common Safety Indicators, Denmark 2014, Version 1, Validated (R11).” Accessed December 26, 2015. <https://erail.era.europa.eu/csi-data.aspx?country=5&year=2014&public=1>.
- “Common Safety Indicators Reported by the National Safety Authorities - R11 - National Value of Preventing a Fatality - Denmark 2006-2014.” Accessed October 1, 2013. <http://erail.era.europa.eu/safety-indicators.aspx>.
- “Common Safety Indicators Reported by the National Safety Authorities - R16 - Fall Back Value of Preventing a Fatality - Denmark 2006-2014.” Accessed October 1, 2013. <http://erail.era.europa.eu/safety-indicators.aspx>.
- “Complicated or Complex - Knowing the Difference Is Important.” *Sparksforchange*. Accessed August 24, 2015. <http://learningforsustainability.net/sparksforchange/complicated-or-complex-knowing-the-difference-is-important-for-the-management-of-adaptive-systems/>.
- Conviction in the case of hacking of CSC, (municipal court of Frederiksberg 2014).
- “Core PKI Services: Authentication, Integrity, and Confidentiality.” Accessed December 16, 2015. <https://technet.microsoft.com/en-us/library/cc700808.aspx?f=255&MSPPError=-2147217396>.
- Cousins, Ben. “Weapons of Mass Disruption.” presented at the Weapons of Mass Disruption: Creating The Drowning, GDC ’13 (Game Developers Conference), March 29, 2013. <http://www.gdcvault.com/play/1017751/Weapons-of-Mass-Disruption-Creating>.
- COWI, and Vejdirektoratet. “Trafikøkonomiske Enhedspriser for uheld - Alternative metoder til opgørelse af Velfærdstabet (Arbejdsnotat),” January 2002.
- “Cryptohaze.com • View Topic - CUDA Multiforcer 0.7 Source.” Accessed June 30, 2015. <http://www.cryptohaze.com/forum/viewtopic.php?f=4&t=64>.
- “CUDA C Programming Guide.” Concept. Accessed December 18, 2015. <http://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html#warp-shuffle-functions>.
- “CUDA Occupancy Calculator Helps Pick Optimal Thread Block Size - NVIDIA Developer Forums.” Accessed December 26, 2015. <https://devtalk.nvidia.com/default/topic/368105/cuda-occupancy-calculator-helps-pick-optimal-thread-block-size/>.
- Dang, Quynh H. “Secure Hash Standard (SHA-1) NIST FIPS 180-4.” National Institute of Standards and Technology, July 2015. <http://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.180-4.pdf>.
- Danish Agency for Digitisation. “Certificate Policy for OCES Employee Certificates (Public Certificates for Electronic Services),” August 2005. https://www.nemid.nu/dk-da/digital_signatur/oces-standarden/oces-certifikatpolitikker/OCES_employee_certificates_version_4.pdf.
- — —. “Certificate Policy for OCES Personal Certificates (Public Certificates for Electronic Services),” September 2009. https://www.nemid.nu/dk-da/digital_signatur/oces-standarden/oces-certifikatpolitikker/POCES_Certifikatpolitik_version_4_Eng.pdf.
- Danish Agency for Digitisation, and Nikolas Triantafyllidis. “Certifikatpolitik for OCES-Personcertifikater (Offentlige Certifikater Til Elektronisk Service) Version 4,” September 2009. https://www.nemid.nu/dk-da/digital_signatur/oces-standarden/oces-certifikatpolitikker/POCES_Certifikatpolitik_version_4.pdf.



- Danish Ministry of Transport, and COWI. "Rapport om værdisætning af transportens eksterne omkostninger." Danish Ministry of Transport, June 2010. <http://www.trm.dk/~media/Files/Publication/2010/Rapport%20om%20v%C3%A6rdi%C3%A6tning%20af%20transportens%20eksterne%20omkostninger.pdf>.
- "DANMARK (DENMARK) : Trusted List." Accessed August 14, 2015. <http://www.digst.dk/~media/Files/Loesninger-og-infrastruktur/NemID/HumanReadableTldkxml.pdf>.
- "DER Encoding of ASN.1 Types (Windows)." Accessed December 7, 2015. [https://msdn.microsoft.com/en-us/library/windows/desktop/bb648640\(v=vs.85\).aspx](https://msdn.microsoft.com/en-us/library/windows/desktop/bb648640(v=vs.85).aspx).
- Dhungle, Prithula, Di Wu, Brad Schonhorst, and Keith W. Ross. "A Measurement Study of Attacks on BitTorrent Leechers." In *IPTPS*, 8:7–7, 2008. <http://www.iptps.org/%5C/papers-2008/47.pdf>.
- Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on a Community Framework for Electronic Signatures, 1999. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31999L0093:EN:HTML>.
- "Djpholy/libgfshare." *GitHub*. Accessed January 11, 2016. <https://github.com/djpholy/libgfshare>.
- Doyle, Arthur Conan, and Charles Henry Malcolm Kerr. *The Sign of Four*. London: Spencer Blackett, 1890.
- Dublin, Louis I., and Alfred J. Lotka. *The Money Value of a Man*. New York vols. Ronald Press, 1930.
- Durumeric, Zakir, Mathias Payer, Vern Paxson, James Kasten, David Adrian, J. Alex Halderman, Michael Bailey, et al. "The Matter of Heartbleed," 475–88. ACM Press, 2014. doi:10.1145/2663716.2663755.
- "EC2 Instance Pricing – Amazon Web Services (AWS)." *Amazon Web Services, Inc.* Accessed January 7, 2016. <http://aws.amazon.com/ec2/pricing/>.
- "Elpriser – Se de Aktuelle Elpriser Hos DONG Energy." Accessed January 6, 2016. <https://www.dongenergy.dk/privat/produkter-og-priser/el>.
- "ERTMS EuroRadio Conformance Requirements - SUBSET-092-1." Accessed December 3, 2015. <http://www.era.europa.eu/Document-Register/Pages/ERTMS-EuroRadio-Conformance-Requirements.aspx>.
- "ERTMSFormalSpecs InnoInstaller5/whatsnew.htm." *GitHub*. Accessed December 16, 2015. <https://github.com/ERTMSSolutions/ERTMSFormalSpecs>.
- "ERTMS Signaling Levels | ERTMS." Accessed December 6, 2015. http://www.ertms.net/?page_id=42.
- "ERTMS Solutions | ERTMSFormalSpecs - Open Source - ERTMS Solutions." Accessed October 12, 2015. <https://www.ertmsolutions.com/products/ertmsformalspecs-open-source/>.
- "ETCS Software Error Led to Lötschberg Derailment." *Railway Gazette*. Accessed December 6, 2015. <http://www.railwaygazette.com/news/single-view/view/etcs-software-error-led-to-loetschberg-derailment.html>.
- EUNET / European Commission. "Socio-Economic and Spatial Impacts of Transport." 4th RTD Framework Programme, March 2001. <http://www.transport-research.info/sites/default/files/project/documents/eunet.pdf>.
- European Railway Agency. "ERTMS Euroradio Test Cases Safety Layer - SUBSET-092-2." Accessed December 3, 2015. <http://www.era.europa.eu/Document-Register/Pages/Set-2-ERTMS-Euroradio-Test-cases-Safety-Layer.aspx>.
- European Railway Agency Corporate Management and Evaluation. "FW: Information Request Form - Nielsen (Dec 2)," December 2, 2015.
- . "FW: Information Request Form - Nielsen (Dec 3)," December 3, 2015.
- "EuroRadio FIS - SUBSET-037." Accessed December 3, 2015. <http://www.era.europa.eu/Document-Register/Pages/Set-2-EuroRadio-FIS.aspx>.
- "Expert Advice: Encryption 101 -- Triple DES Explained." *SearchSecurity*. Accessed December 5, 2015. <http://searchsecurity.techtarget.com/tip/Expert-advice-Encryption-101-Triple-DES-explained>.
- "Facebook's WhatsApp Hits 900 Million Users, Aims for 1 Billion." *USA TODAY*. Accessed January 2, 2016. <http://www.usatoday.com/story/tech/2015/09/04/whatsapp-facebook-900-million-mark-zuckerberg-jan-koum-messenger/71704760/>.
- "Fatal Plane Crash Rates by Model." Accessed January 3, 2016. http://www.airsafe.com/events/models/rate_mod.htm.
- "FRANCE (FRANCE) : Trusted List." Accessed December 23, 2015. <http://references.modernisation.gouv.fr/sites/default/files/TSL-FR.xml.pdf>.
- "FRB: How Much U.S. Currency Is in Circulation?" Accessed January 7, 2016. http://www.federalreserve.gov/faqs/currency_12773.htm.



- Freibott, Bernd. "Sustainable Safety Management: Incident Management as a Cornerstone for a Successful Safety Culture," 2012.
<https://books.google.com/books?hl=en&lr=&id=oFBX074a04cC&oi=fnd&pg=PA257>.
- "Further Safety Measures Follow Santiago de Compostela Crash." *Railway Gazette*. Accessed December 5, 2015.
<http://www.railwaygazette.com/news/policy/single-view/view/further-safety-measures-follow-santiago-crash.html>.
- German Federal Bureau of Aircraft Accidents Investigation. "Überlingen Mid-Air Collision Investigation Report." German Federal Bureau of Aircraft Accidents Investigation, May 2004. http://www.bfu-web.de/EN/Publications/Investigation%20Report/2002/Report_02_AX001-1-2_Ueberlingen_Report.pdf?__blob=publicationFile.
- "Google/end-to-End." *GitHub*. Accessed July 29, 2015. <https://github.com/google/end-to-end>.
- "Google/end-to-End - Source Code Search for SHA." *GitHub*. Accessed August 22, 2015.
<https://github.com/google/end-to-end>.
- "Google/end-to-End Userid.js." *GitHub*. Accessed December 21, 2015. <https://github.com/google/end-to-end/blob/7fa39bb1cce553ce39c42af5eebb7aac46d2fe1d/src/javascript/crypto/e2e/openpgp/packet/userid.js>.
- Great Britain. Health and Safety Executive. *Reducing Risks, Protecting People*. Sudbury: HSE Books, 2001.
- Grunthal, Aaron. "Efficient Indexing of the BitTorrent Distributed Hash Table." *arXiv Preprint arXiv:1009.3681*, 2010. <http://arxiv.org/abs/1009.3681>.
- Guido, B., D. Joan, P. Michaël, V. A. Gilles, and V. K. Ronny. "Keccak Implementation Overview," 2011.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.361.7964&rep=rep1&type=pdf>.
- Gürgens, S., and C. Rudolph. "Security Analysis of Efficient (Un-) Fair Non-Repudiation Protocols." *Formal Aspects of Computing* 17, no. 3 (March 22, 2005): 260–76. doi:10.1007/s00165-004-0055-4.
- "HackerOne: Vulnerability Coordination and Bug Bounty Platform." Accessed July 14, 2015.
<https://hackerone.com/>.
- Harrison, David. "Index of BitTorrent Enhancement Proposals." *Index of BitTorrent Enhancement Proposals*. Accessed July 13, 2015. http://www.bittorrent.org/beps/bep_0000.html.
- Hase, Klaus-Rüdiger. "'Open Proof' for Railway Safety Software - A Potential Way-Out of Vendor Lock-in Advancing to Standardization, Transparency, and Software Security." In *FORMS/FORMAT 2010*, edited by Eckehard Schnieder and Geza Tarnai, 5–38. Springer Berlin Heidelberg, 2011.
http://link.springer.com/chapter/10.1007/978-3-642-14261-1_2.
- Heinrich, Herbert William. *Industrial Accident Prevention: A Scientific Approach*. McGraw-Hill book Company, Incorporated, 1931.
- "How to Verify the Authenticity of Manually Downloaded Apple Software Updates - Apple Support." Accessed December 14, 2015. <https://support.apple.com/en-us/HT202369>.
- Hultkrantz, Lars, and Mikael Svensson. "The Value of a Statistical Life in Sweden: A Review of the Empirical Literature." *Health Policy* 108, no. 2–3 (December 2012): 302–10. doi:10.1016/j.healthpol.2012.09.007.
- Imperial Chemical Industries, Ltd, Chemical Industries Association, and Chemical Industry Safety & Health Council. *A Guide to Hazard and Operability Studies*. London: Chemical Industry Safety and Health Council of the Chemical Industries Association, 1977.
- "Implementation Guidelines for NemID (OCES) Version 2.1." Accessed November 15, 2015.
<http://www.nets.eu/dk-da/Service/kundeservice/nemid-tu/tjenesteudbyderpakkeJS/Documents/NemID%20Integration%20-%20OCES.pdf>.
- Institute for Defense Analyses. "Open Source Software (OSS/FLOSS) and Security International Workshop on Free/Open Source Software Technologies Riyadh, Saudi Arabia." September 22, 2011.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.169.8525&rep=rep1&type=pdf>.
- "Intent to Deprecate: SHA-1 Certificates - Google Groups," August 20, 2014.
<https://groups.google.com/a/chromium.org/forum/#!msg/security-dev/2-R4XziFc7A/NDI8cOwMGRQJ>.
- International Electrotechnical Commission. "IEC 61508 Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems," June 1, 2010.
- "iOS 5 and iOS 6: List of Available Trusted Root Certificates - Apple Support." Accessed January 4, 2016.
<https://support.apple.com/en-us/HT201388>.
- "ÍSLAND (ICELAND) : Trusted List." Accessed December 23, 2015. <http://docplayer.net/3846400-Island-iceland-trusted-list.html>.
- ISO. "ISO 32000-1:2008: Portable Document Format," July 2008.
- "ITALIA (ITALY) : Trusted List." Accessed December 23, 2015.
https://applicazioni.cnipa.gov.it/TSL/IT_TSL_CNS.pdf.



- Itoh, Kouichi, Tetsuya Izu, Wakaha Ogata, Takeshi Shimoyama, and Masahiko Takenaka. "Forgery Attacks on Time-Stamp, Signed PDF and X.509 Certificate." *IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences* 92 (2009): 67–75. doi:10.1587/transfun.E92.A.67.
- "Jernbanelov - Retsinformation.dk." Accessed December 26, 2015. <https://www.retsinformation.dk/Forms/R0710.aspx?id=170457>.
- Joint Council on Transit Wireless Communications. "Positive Train Control White Paper." Transit Technology, May 2012. http://transitwireless.org/wp-content/uploads/2012/05/PTC_whitepaper_may2012_ver2.pdf.
- Jørgensen, Morten Lisborg. "Analysis and Enhancement of Safety Critical Communication for Railway Systems." Aalborg university, Department of Electronic Systems, 2008. <http://projekter.aau.dk/projekter/da/studentthesis/analysis-and-enhancement-of-safetycritical-communication-for-railway-systems%28cc87b468-6c18-4ed5-ab7d-fb9c9b1d26e6%29.html>.
- Jovicic, Dragan. "ERA Guide for Application of the Common Safety Methods on Risk Assessment." Accessed July 1, 2015. <http://www.era.europa.eu/Document-Register/Pages/guide-for-application-common-safety-method-risk-assessment.aspx>.
- Karpman, Pierre, Thomas Peyrin, and Marc Stevens. "Practical Free-Start Collision Attacks on 76-Step SHA-1," 2015. <https://eprint.iacr.org/2015/530>.
- Kidholm, Kristian, Odense Universitet, and Center for Helsetjenesteforskning og Socialpolitik. "Estimation af betalingsvilje for forebyggelse af personskader ved trafikulykker." Odense Universitet, Det Samfundsvidenskabelige Fakultet, 1995.
- Klima, Vlastimil. "Finding MD5 Collisions-a Toy For a Notebook." *IACR Cryptology ePrint Archive* 2005 (2005): 75.
- Klutke, G., P.C. Kiessler, and M.A. Wortman. "A Critical Look at the Bathtub Curve." *IEEE Transactions on Reliability* 52, no. 1 (March 2003): 125–29. doi:10.1109/TR.2002.804492.
- "Known Logs - Certificate Transparency." Accessed October 26, 2015. <https://www.certificate-transparency.org/known-logs>.
- Kong, Jie, Wandong Cai, Lei Wang, and Qiushi Zhao. "A Study of Pollution on BitTorrent." In *Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on*, 3:118–22. IEEE, 2010. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5452055.
- KPMG IT Advisory. "ERTMS IT Security Threat Identification, Risk Analysis and Recommendations PUBLIC VERSION," April 2013. http://ertms.be/pdf/IT_Security_Threat_identification.pdf.
- Lai, Kevin, Michal Feldman, Ion Stoica, and John Chuang. "Incentives for Cooperation in Peer-to-Peer Networks." In *Workshop on Economics of Peer-to-Peer Systems*, 1243–48, 2003. https://www.gninet.org/sites/default/files/incentives-for-cooperation-in_0.pdf.
- Lambert, Craig. "Disruptive Genius." *Harvard Magazine*, July 2014. <http://harvardmagazine.com/2014/07/disruptive-genius>.
- "Language Solutions." *NVIDIA Developer*. Accessed December 5, 2015. <https://developer.nvidia.com/language-solutions>.
- "LATVIJA (LATVIA) : Trusted List." Accessed December 23, 2015. <http://www.dvi.gov.lv/en/wp-content/uploads/TSL/tsl-lv-6.pdf>.
- "Legal Uses For BitTorrent: You'd Be Surprised." *MakeUseOf*. Accessed December 27, 2015. <http://www.makeuseof.com/tag/8-legal-uses-for-bittorrent-you-d-be-surprised/>.
- "Lists of Available Trusted Root Certificates in OS X - Apple Support." Accessed December 14, 2015. <https://support.apple.com/en-us/HT202858>.
- Loewenstern, Andrew, and Arvid Nordberg. "DHT Protocol," January 31, 2008. http://www.bittorrent.org/beps/bep_0005.html.
- Lov Om Elektroniske Signaturer (Act No. 417 of 31 May 2000 on Electronic Signatures). Act No. 417 of 31 May 2000 on Electronic Signatures*, 2000. <https://www.retsinformation.dk/forms/r0710.aspx?id=6193#>.
- "Luxembourg (Luxembourg): Trusted List." Accessed December 23, 2015. <http://www.portail-qualite.public.lu/fr/actualites/confiance-numerique/2013/nouvelle-trusted-list-18-10-2013/TSL-PDF.pdf>.
- "MALTA (MALTA) : Trusted List." Accessed December 23, 2015. https://www.mca.org.mt/tsl/MT_TSL.pdf.
- Marc, Stevens. "Cryptanalysis of MD5 & SHA-1." Accessed June 30, 2015. <http://2012.sharcs.org/slides/stevens.pdf>.
- Mária Franeková, Karol Rástočný. "Safety Analysis of Cryptography Mechanisms Used in GSM for Railway." *Annals of Faculty Engineering Hunedoara - International Journal of Engineering* IX, no. 1 (2011): 207–12.
- Martin, Scott. *BitTorrent Network*, February 15, 2014. https://commons.wikimedia.org/wiki/File:BitTorrent_network.svg.



- McVittie, Simon. "Theory Used by Libgfsahre," April 23, 2006. <http://www.digitalscurf.org/files/libgfsahre/theory.pdf>.
- "Measurements of Hash Functions, Indexed by Machine." Accessed October 16, 2015. <http://bench.cr.yt.to/results-hash.html>.
- Merkle, Ralph C. "A Certified Digital Signature." In *Advances in Cryptology — CRYPTO' 89 Proceedings*, edited by Gilles Brassard, 218–38. Lecture Notes in Computer Science 435. Springer New York, 1989. http://link.springer.com.proxy.findit.dtu.dk/chapter/10.1007/0-387-34805-0_21.
- "Microsoft and Facebook Launch Internet Bug Bounty Program." *The Next Web*. Accessed July 21, 2015. <http://thenextweb.com/insider/2013/11/06/microsoft-facebook-sponsor-internet-bug-bounty-program-offer-cash-hacking-internet-stack/>.
- "Microsoft Word - MD5 Collisions Whitepaper.doc - wp.MD5_Collisions.en_us.pdf." Accessed December 20, 2015. https://ad-pdf.s3.amazonaws.com/papers/wp.MD5_Collisions.en_us.pdf.
- "Netscape Bugs Bounty," October 10, 1995. <https://web.archive.org/web/19970501041756/www101.netscape.com/newsref/pr/newsrelease48.html>.
- Nordber, Arvid. "uTorrent Transport Protocol." Accessed January 10, 2016. http://bittorrent.org/beps/bep_0029.html.
- "NVCC :: CUDA Toolkit Documentation." Accessed December 17, 2015. <http://docs.nvidia.com/cuda/cuda-compiler-driver-nvcc/#axzz3uZw6Czwc>.
- NVIDIA. "Kepler Compute Architecture Whitepaper." Accessed October 5, 2015. <http://international.download.nvidia.com/pdf/kepler/NVIDIA-Kepler-GK110-GK210-Architecture-Whitepaper.pdf>.
- "Offline Key Management FIS - SUBSET-038." Accessed December 3, 2015. <http://www.era.europa.eu/Document-Register/Pages/Set-2-Offline-key-management-FIS.aspx>.
- "Open Data Catalog | The World Bank." Accessed January 7, 2016. <http://datacatalog.worldbank.org/>.
- "OpenSSL Audit." Accessed July 13, 2015. <https://cryptoservices.github.io/openssl/2015/03/09/openssl-audit.html>.
- Park, Sunwoo, Changbin Lee, Kwangwoo Lee, Jeeyeon Kim, Youngsook Lee, and Dongho Won. "Security Analysis on DigitalSignature Function Implemented in PDF Software." In *Future Generation Information Technology*, edited by Tai-hoon Kim, Hojjat Adeli, Dominik Slezak, Frode Eika Sandnes, Xiaofeng Song, Kyo-il Chung, and Kirk P. Arnett, 327–34. Lecture Notes in Computer Science 7105. Springer Berlin Heidelberg, 2011. http://link.springer.com.proxy.findit.dtu.dk/chapter/10.1007/978-3-642-27142-7_38.
- "Paypal: Quarterly Net Revenue 2015 | Statistic." *Statista*. Accessed December 27, 2015. <http://www.statista.com/statistics/218517/paypals-net-revenue-per-quarter/>.
- Pfleeger, Charles P., and Shari Lawrence Pfleeger. *Security in Computing*. 4th ed. Upper Saddle River, NJ: Prentice Hall, 2007.
- Pgi-Kepler-Block-Diagram.png (PNG Image, 1152 × 864 Pixels) - Scaled (79%)*. Accessed December 5, 2015. <https://www.pgroup.com/images/insider/pgi-kepler-block-diagram.png>.
- "POLSKA (POLAND) : Trusted List." Accessed December 23, 2015. https://www.nccert.pl/tsl/PL_TSL.pdf.
- Pratt, John W., and Richard J. Zeckhauser. "Willingness to Pay and the Distribution of Risk and Wealth." *Journal of Political Economy* 104, no. 4 (August 1, 1996): 747–63. doi:10.2307/2138884.
- "Product Details." *Amazon Web Services, Inc.* Accessed January 7, 2016. <http://aws.amazon.com/ec2/details/>.
- "Project Zero: Announcing Project Zero." Accessed August 24, 2015. <http://googleprojectzero.blogspot.com/2014/07/announcing-project-zero.html>.
- Quynh, Dang. "Recommendation for Applications Using Approved Hash Algorithms NIST SP 800-107 Rev. 1." *National Institute of Standards and Technology (NIST) Special Publication* 800, no. 107 (August 2012): 108.
- Rail Safety & Standards Board. "T430 Assessment of the Value for Preventing a Fatality Phase 1." Accessed April 13, 2013. <http://www.rssb.co.uk/SiteCollectionDocuments/pdf/reports/research/T430%20Assessment%20of%20the%20Value%20for%20Preventing%20a%20Fatality%20Phase%201.pdf>.
- . "T430 Assessment of the Value for Preventing a Fatality Phase 1." Accessed April 13, 2013. <http://www.rssb.co.uk/SiteCollectionDocuments/pdf/reports/research/T430%20Assessment%20of%20the%20Value%20for%20Preventing%20a%20Fatality%20Phase%201.pdf>.
- "Railway Applications - Communication, Signalling and Processing Systems - Safety-Related Communication in Transmission Systems - EN 50159," n.d.
- Reason, J. *Managing the Risks of Organizational Accidents*. Ashgate, 1997.



- “Regler - Om NemID - NemID (verified January 11-2016).” Accessed January 11, 2016.
https://www.nemid.nu/dk-da/om_nemid/regler/.
- “Researchers Hijack Printer Using Malicious Firmware Update.” Accessed December 28, 2015.
<http://www.eweek.com/c/a/Security/Researchers-Hijack-Printer-Using-Malicious-Firmware-Update-856123>.
- Reynolds, D. J. “The Cost of Road Accidents.” *Journal of the Royal Statistical Society. Series A (General)* 119, no. 4 (January 1, 1956): 393–408. doi:10.2307/2342577.
- “RFC 2459 X509 Cert - Obsolete.” Accessed December 7, 2015. <https://www.rfc-editor.org/rfc/rfc2459.txt>.
- “ROMÂNIA (ROMANIA) : Trusted List.” Accessed December 23, 2015. http://www.mcsi.ro/Minister/Domenii-de-activitate-ale-MCSI/Tehnologia-Informatiei/Servicii-electronice/Semnatura-electronica/ROMANIA_TrustedList-v7-pdf.
- Roth, Thomas. “Cracking Passwords In The Cloud: Amazon’s New EC2 GPU Instances | Stacksmashing.net,” December 31, 2011.
<https://web.archive.org/web/20111231044601/http://stacksmashing.net/2010/11/15/cracking-in-the-cloud-amazons-new-ec2-gpu-instances>.
- “S2K Uses Small C/bytcount, Inconsistent Suite of S2K-KDF-SHA1 (160b) and AES-256 · Issue #139 · Google/end-to-End.” *GitHub*. Accessed August 22, 2015. <https://github.com/google/end-to-end/issues/139>.
- Sandvine, Intelligent broadband networks. “Sandvine Global Internet Phenomena Report - 2H 2014 - 2h-2014-Global-Internet-Phenomena-Report.pdf.” *Sandvine Global Internet Phenomena Report - 2H 2014 - 2h-2014-Global-Internet-Phenomena-Report*. Accessed July 13, 2015.
<https://www.sandvine.com/downloads/general/global-internet-phenomena/2014/2h-2014-global-internet-phenomena-report.pdf>.
- Sargut, Gökçe, and Rita McGrath. “Learning to Live with Complexity.” *Harvard Business Review*, September 2011.
<https://hbr.org/2011/09/learning-to-live-with-complexity>.
- Schelling, Thomas C. *Choice and Consequence*. Harvard University Press, 1984.
- Schneier, Bruce. “When Will We See Collisions for SHA-1? - Schneier on Security.” Accessed June 30, 2015.
https://www.schneier.com/blog/archives/2012/10/when_will_we_se.html.
- Schweizerische Eidgenossenschaft. “Frutigen ERTMS derailment report (Schlussbericht der Unfalluntersuchungsstelle Bahnen und Schiffe über die Entgleisung von Güterzug 43647 der BLS AG auf der Weiche 34 (Einfahrt Lötschberg-Basistrecke) vom Dienstag, 16. Oktober 2007 in Frutigen),” June 23, 2008. http://www.sust.admin.ch/pdfs/BS//pdf/07101601_SB.pdf.
- “Set of Specifications # 2 (ETCS Baseline 3 and GSM-R Baseline 0).” Accessed October 27, 2015.
<http://www.era.europa.eu/Core-Activities/ERTMS/Pages/Set-of-specifications-2.aspx>.
- “SHA1 Deprecation Policy - Windows PKI Blog - Site Home - TechNet Blogs,” November 12, 2013.
<http://blogs.technet.com/b/pki/archive/2013/11/12/sha1-deprecation-policy.aspx>.
- “Sha 1 - Why Initialize SHA1 with Specific Buffer? - Cryptography Stack Exchange.” Accessed January 11, 2016.
<http://crypto.stackexchange.com/questions/10829/why-initialize-sha1-with-specific-buffer>.
- Shaw, David, Lutz Donnerhacke, Rodney Thayer, Hal Finney, and Jon Callas. “OpenPGP Message Format - RFC 4880.” Accessed July 28, 2015. <https://tools.ietf.org/html/rfc4880>.
- Shogren, Jason F., Seung Y. Shin, Dermot J. Hayes, and James B. Kliebenstein. “Resolving Differences in Willingness to Pay and Willingness to Accept.” *The American Economic Review* 84, no. 1 (March 1, 1994): 255–70. doi:10.2307/2117981.
- “Simple Linux Utility for Resource Management.” Accessed December 21, 2015. <http://slurm.schedmd.com/>.
- “Slurm Job Scheduler.” Accessed December 21, 2015. <https://deic.sdu.dk/documentation/slurm>.
- Stafford E., Tavares. *On the Design of S-Boxes - Advances in Cryptology*. Springer-Verlag New York, Inc., n.d.
- Statistics Denmark. “Traffic Accidents with Injuries.” Accessed January 3, 2016.
<http://www.statistikbanken.dk/statbank5a/SelectVarVal/Define.asp?MainTable=UHELD4&PLanguage=0&PXSID=0&wsid=cftree>.
- Stephan. “Making End-to-End Encryption Easier to Use.” *Google Online Security Blog*, June 3, 2014.
<http://googleonlinesecurity.blogspot.com/2014/06/making-end-to-end-encryption-easier-to.html>.
- Stevens, Marc. “Cryptanalysis of MD5 & SHA-1.” Accessed June 30, 2015.
<http://2012.sharcs.org/slides/stevens.pdf>.
- Stevens, Marc, Pierre Karpman, and Thomas Peyrin. “Freestart Collision on Full SHA-1,” 2015.
<https://eprint.iacr.org/2015/967>.
- Stevens, Marc, Alexander Sotirov, Jacob Appelbaum, Arjen Lenstra, David Molnar, Dag Arne Osvik, and Benne de Weger. “Short Chosen-Prefix Collisions for MD5 and the Creation of a Rogue CA Certificate.” In *Advances in Cryptology - CRYPTO 2009*, edited by Shai Halevi, 55–69. Lecture Notes in Computer Science 5677.



- Springer Berlin Heidelberg, 2009. http://link.springer.com.proxy.findit.dtu.dk/chapter/10.1007/978-3-642-03356-8_4.
- “STM FFFIS Safe Link Layer - SUBSET 057.” Accessed October 27, 2015. <http://www.era.europa.eu/Document-Register/Pages/STM-FFFIS-Safe-link-Layer.aspx>.
- Svensson, Peter. “Consumer Groups Ask FCC to Fine Comcast.” *Msnbc.com*. Accessed December 26, 2015. http://www.nbcnews.com/id/21579686/ns/technology_and_science/t/consumer-groups-ask-fcc-fine-comcast/.
- “TERENA> News> TCS Certificate Service Responds to SHA Security Update.” Accessed September 7, 2015. https://www.terena.org/news/fullstory.php?news_id=3733.
- “Terms of Service - HackerOne.” Accessed August 24, 2015. <https://hackerone.com/terms>.
- “The 5 Most Popular BitTorrent Trackers.” *TorrentFreak*. Accessed July 20, 2015. <https://torrentfreak.com/5-most-popular-bittorrent-trackers-070924/>.
- The Brønnøysund Register Centre (Norwegian Business Registry). “Change of Business Enterprise Name STARTUP 629 14 AS to NASSA MIDCO AS.” Accessed January 4, 2016. http://w2.brreg.no/kunngjoring/hent_en.jsp?kid=20140000068129&sokeverdi=913111990&spraak=en.
- . “NASSA MIDCO AS Organization Number: 913 111 990.” Accessed January 4, 2016. http://w2.brreg.no/kunngjoring/hent_nr.jsp?orgnr=913111990&spraak=en.
- The European parliament and council. *Directive 2013/40/EU (Cybercrime)*. 32013L0040, 2013. <http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32013L0040>.
- “The Hidden Costs of Heartbleed.” *CloudFlare*. Accessed July 28, 2015. <http://blog.cloudflare.com/the-hard-costs-of-heartbleed/>.
- “The History of Bug Bounty Programs.” Accessed July 14, 2015. <https://www.crowdcurity.com/blog/the-history-of-bug-bounty-programs>.
- The Internet Society. “RFC 3280 - Internet X.509 Public Key Infrastructure.” Accessed October 4, 2015. <https://www.ietf.org/rfc/rfc3280.txt>.
- Theory Team. “BitTorrentPeerExchangeConventions - Theory.org Wiki,” April 228AD. <https://wiki.theory.org/BitTorrentPeerExchangeConventions>.
- “The Shappening.” Accessed October 16, 2015. <https://sites.google.com/site/itstheshappening/>.
- “THIRD (FINAL) ROUND SHA-3 CANDIDATES- NIST.gov - Computer Security Division - Computer Security Resource Center.” Accessed January 11, 2016. http://csrc.nist.gov/groups/ST/hash/sha-3/Round3/submissions_rnd3.html.
- “Top 10 Most Popular Torrent Sites of 2014.” *TorrentFreak*. Accessed July 20, 2015. <https://torrentfreak.com/top-10-popular-torrent-sites-2014-140104/>.
- “Top 10 Most Popular Torrent Sites of 2015.” *TorrentFreak*. Accessed July 20, 2015. <https://torrentfreak.com/top-popular-torrent-sites-2015-150104/>.
- “TOP500 Supercomputer Sites | 267.” Accessed October 16, 2015. <http://www.top500.org/system/178547>.
- “Torcache - Torrent Cache.” Accessed December 28, 2015. <http://torcache.net/api>.
- Torkington, Nat. “HBO Attacking BitTorrent - O’Reilly Radar.” Accessed December 26, 2015. <http://radar.oreilly.com/2005/10/hbo-attacking-bittorrent.html>.
- Transcript of hacker case, (municipal court of Frederiksberg 2014).
- Treich, Nicolas. “The Value of a Statistical Life under Ambiguity Aversion.” *Journal of Environmental Economics and Management* 59, no. 1 (January 2010): 15–26. doi:10.1016/j.jeem.2009.05.001.
- “Trusted Service List - Dansk.” Accessed August 14, 2015. <http://www.digst.dk/Loesninger-og-infrastruktur/NemID/Internationalt-samarbejde/Trusted-List-Danmark>.
- “UNITED KINGDOM (UNITED KINGDOM) : Trusted List.” Accessed December 23, 2015. http://www.tscheme.org/UK_TSL/TSL-UK0006signed.pdf.
- US Department of Commerce, NIST. “NIST Selects Winner of Secure Hash Algorithm (SHA-3) Competition.” Accessed December 20, 2015. <http://www.nist.gov/itl/csd/sha-100212.cfm>.
- Virk (Danish Business Registry). “NASSA A/S (CVR 34903360).” Text. *Data*. Accessed January 4, 2016. <https://datacvr.virk.dk/data/visenhed?enhedstype=virksomhed&id=34903360&type=Alle&language=en-gb>.
- . “NETS A/S (CVR 20016175).” Text. *Data*. Accessed January 4, 2016. <https://datacvr.virk.dk/data/visenhed?enhedstype=virksomhed&id=20016175&type=Alle&language=en-gb>.
- . “NETS DANID A/S (CVR 30808460).” Text. *Data*. Accessed January 4, 2016. <https://datacvr.virk.dk/data/visenhed?enhedstype=virksomhed&id=30808460&type=Alle&language=en-gb>.



- . “NETS HOLDING A/S (CVR 27225993).” Text. *Data*. Accessed January 4, 2016. <https://datacvr.virk.dk/data/visenhed?enhedstype=virksomhed&id=27225993&type=Alle&language=en-gb>.
- Vishnumurthy, Vivek, Sangeeth Chandrakumar, and Emin Gun Sirer. “Karma: A Secure Economic Framework for Peer-to-Peer Resource Sharing.” In *Workshop on Economics of Peer-to-Peer Systems*, Vol. 35, 2003. <http://kayapo.tribler.org/trac/raw-attachment/wiki/ExistingReputationSystems/KARMA,%20A%20Secure%20Economic%20Framework%20for%20Peer-to-Peer%20Resource%20Sharing.pdf>.
- “Vulnerability Disclosure Policy | Vulnerability Analysis | The CERT Division.” Accessed August 21, 2015. <https://www.cert.org/vulnerability-analysis/vul-disclosure.cfm>.
- Vuze Team. “Peer Exchange - VuzeWiki,” May 12, 2010. http://wiki.vuze.com/w/Peer_Exchange.
- . *Peers with Pieces*, June 3, 2010. <https://wiki.vuze.com/w/File:PeersWithRandomPieces.png>.
- . “Vuze Open-Source BitTorrent Client Documentation,” March 2, 2010. https://wiki.vuze.com/w/Torrent_Piece_Size.
- Wang, Jian, Xiaoming Hu, Yinchun Yang, and Xiumei Wu. “A Misbehavior Resilient Cipherblock Trading Protocol in BitTorrent-like Networks.” In *Computer Science and Network Technology (ICCSNT), 2011 International Conference on*, 2:885–88. IEEE, 2011. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6182103.
- Weisstein, Eric W. “Birthday Problem.” Text. Accessed December 27, 2015. <http://mathworld.wolfram.com/BirthdayProblem.html>.
- Weste, Neil H. E., and David Money Harris. *Integrated Circuit Design*. 4. ed., global ed. Boston, Mass.: Pearson, 2011.
- Wheeler, David A. “Secure Software Design & Programming - Formal Methods.” May 5, 2015. <http://www.dwheeler.com/secure-class/presentations/Secure-Software-10-Formal-Methods.ppt>.
- Willumsen, Eva, Mads Paabøl Jensen, and Per Skrumdsager Hansen. “Nye Værdier for Transportens Eksterne Omkostninger,” 2010. http://www.trafikdage.dk/papers_2010/374_EvaWillumsen.pdf.
- Yang, Beverly, and Hector Garcia-Molina. “PPay: Micropayments for Peer-to-Peer Systems.” In *Proceedings of the 10th ACM Conference on Computer and Communications Security*, 300–310. ACM, 2003. <http://dl.acm.org/citation.cfm?id=948150>.
- “Zero Day Initiative - Disclosure Policy.” Accessed August 21, 2015. http://www.zerodayinitiative.com/advisories/disclosure_policy/.
- Zhang, Hao, Yonggang Wen, Haiyong Xie, and Nenghai Yu. *Distributed Hash Table*. SpringerBriefs in Computer Science. New York, NY: Springer New York, 2013. <http://link.springer.com/10.1007/978-1-4614-9008-1>.
- “ΚΥΠΡΟΣ/ΚΙΒΙΣ (CYPRUS) : Trusted List.” Accessed December 23, 2015. [http://www.mcw.gov.cy/mcw/dec/dec.nsf/all/0F90C75AEE05D35DC22577E400253132/\\$file/TSL-CY-003-sign.pdf?openelement](http://www.mcw.gov.cy/mcw/dec/dec.nsf/all/0F90C75AEE05D35DC22577E400253132/$file/TSL-CY-003-sign.pdf?openelement).
- “БЪЛГАРИЯ (BULGARIA) : Trusted List.” Accessed December 23, 2015. http://www.crc.bg/files/_en/TSL-BG-CRC-signed.pdf.



19 Abbreviations, technical terms & definitions

19.1 Abbreviations

AES: **A**dvanced **E**ncryption **S**tandard;

ALARP: **A**s **L**ow **A**s **R**easonably **P**racticable;

ATO: **A**utomatic **T**rain **O**peration;

BT: **B**it**T**orrent; a protocol for transferring data in a Peer to Peer network.

CA: **C**ertificate **A**uthority;

CBA: **C**ost **B**enefit **A**nalysis;

CUDA: **C**ompute **U**nified **D**evice **A**rchitecture;

CPU: **C**entral **P**rocessing **U**nit;

CRC: **C**yclic **R**edundancy **C**heck;

DeIC: **D**anish **E**-**I**nfrastructure **C**ooperation;

DES: **D**ata **E**ncryption **S**tandard;

DHT: **D**istributed **H**ash **T**able;

DKK: **D**anish **K**rone;

DDOS: **D**istributed **D**enial **O**f **S**ervice;

DOS: **D**enial **O**f **S**ervice;

DHT: **D**istributed **H**ash **T**able;

E2E: **E**nd **t**o **E**nd;

ENISA: **E**uropean **N**etwork and **I**nformation **S**ecurity **A**gency;

ERA: **E**uropean **R**ail **A**gency;

ERTMS: ETCS + GSM-R;

ETCS: Part of ERTMS;

EU: **E**uropean **U**nion;

EVC: **E**uropean **V**ital **C**omputer;

FFFIS: **F**orm **F**it **F**unctional **I**nterface **S**pecification;

FIS: **F**unctional **I**nterface **S**pecification;

FMEA: **F**ailure **M**ode and **E**ffect **A**nalysis;

GDP: **G**ross **D**omestic **P**roduct;

GPGPU: **G**eneral **P**urpose **G**raphics **P**rocessing **U**nit;

GPU: **G**raphics **P**rocessing **U**nit;

GSM-R: Part of ERTMS;

HAZOP: **H**AZard and **O**Perability analysis;

HDOS: **H**ash **D**enial **o**f **S**ervice

HK: **H**uman **C**apital;

HPC: **H**igh **P**erformance **C**omputing/**C**omputer;

HSM: **H**ardware **S**ecured **M**odule;

HTTP: **H**ypertext **T**ransfer **P**rotocol; an internet information transfer protocol.

HTTPS: **H**TTP**S**ecure; a secured version of **HTTP** using **SSL** or **TLS**.

IEC: **I**nternational **E**lectrotechnical **C**ommission;

IETF: **I**nternet **E**ngineering **T**ask **F**orce;

IP: **I**nternet **P**rotocol;

IPv6: **I**nternet **p**rotocol **v**ersion **6**;

ISO: **I**nternational **O**rganization for **S**tandardization;



JIT: J ust I n T ime;
KPI: K ey P erformance I ndicators;
MCD A: M ulti-Criteria D ecision A nalysis;
MD5: M essage D igest A lgorithm 5;
NIST: N ational I nstitute of S tandards and T echnology;
NSA: N ational S ecurity A uthority;
OCES: O ffentlige C ertifikater til E lektronisk S ervice; (Public Certificates for Electronic Service)
PCI: P eripheral C omponent I nterconnect;
PDF: P ortable D ocument F ormat;
PEX: P eer E XChange;
PTC: P ositive T rain C ontrol;
RP: R evealed P reference
PRNG: P seudo R andom N umber G enerator;
RFC: R elease F or C omments;
RSA: R ivate-S hamir-A dleman cryptosystem;
RSSB: R ail S afety and S tandards B oard;
SDK: S oftware D evelopment K it;
SHA-0: S ecur e H ash A lgorithm 0;
SHA-1: S ecur e H ash A lgorithm 1;
SHA-2: S ecur e H ash A lgorithm 2;
SHA-256: S ecur e H ash A lgorithm 2 (256 bit version);
SHA-3: S ecur e H ash A lgorithm 3;
SIL: S afety I ntegrity L evel;
SLURM: S imple L inux U tility for R esource M anagement;
SMX: S treaming M ultiprocessor;
SSD: S olid S tate D isk;
SSL: S ecur e S ockets L ayer;
SP: S tated P reference
STM: S pecific T ransmission M odule;
TCAS: T raffic C ollision A voidance S ystem;
TCP: T ransmission C ontrol p rotocol;
TLS: T ransport L ayer S ecurity;
TSI: T echnical S pecifications for I nteroperability
TSL: (Trusted Service List);
USD: U nited S tates D ollar;
VPC: V alue of P reventing a C asualty;
VSL: V alue of S tatistical L ife;

19.2 Technical terms and definitions

Complex problem: A problem that does not have just one right solution, but rather a plethora of solutions that each have their benefits and side effects.

“The main difference between complicated and complex systems is that with the former, one can usually predict outcomes by knowing the starting conditions. In a complex system, the same starting conditions can produce different outcomes, depending on the interactions of the elements in the system.” ³³⁶

³³⁶ Sargent and McGrath, “Learning to Live with Complexity.”



*“We can **determine** complicated outcomes. We can only **enable** complex outcomes. We can **specify** complicated systems. We can only **intervene** in complex systems.”³³⁷*

Complicated problem: A computational problem that will take a long time to solve, but has a single right result. This is in contrast to Complex problems (see above).

Disruption: When a new method improves performance exponentially, while the current methods by incumbent companies improve linearly.

Qualifiers for a disruptive innovation:

Cheaper, simpler, smaller, and, frequently, more convenient to use.³³⁸

Result in worse product performance, at least in the near term.³³⁹

Improves a product or service in ways that the market does not expect.³⁴⁰

Non-repudiation: prevents an entity from successfully denying involvement in a previous action.

“Where non-repudiation is indicated, certificate policies commonly include provisions intended to ensure that only one copy of the private key exists, and no party, other than the certificate subject, ever has control of that private key. This is done to protect against repudiation of the signature on the grounds that some party other than the certificate subject might have executed the signature.”³⁴¹

Open proof:

- “Source code, proofs, and required tools: OSS
 - Anyone can examine/critique, improve upon, collaborate with others for improvements
 - Not just software, but what’s proved & tools
 - Example for training, or as useful component
 - Extends OSS idea for high assurance
 - Enables legal collaboration
 - Similar to mathematics field
 - Method for speeding up tech transition
 - Encourage/require government-funded results be open proofs
 - By default – evaluate exceptions
 - Application of “open access” applied broadly
 - See: <http://www.phdcomics.com/comics/archive.php?comicid=1533>
 - Goal: Make supplier identity irrelevant
 - Don’t need everything to be an open proof
 - Examples & building blocks (inc. standards’ API)”
- David A. Wheeler³⁴² and the Institute for Defense Analyses³⁴³. (CC BY-SA 3.0)

Repudiation: “To reject the validity or authority of” see non-repudiation

Researcher: Normally refers to a security researcher; a person that find vulnerabilities a submit them to a Vendor or bug bounty platform

Vendor: A software vendor, the entity responsible for software and hence the point of contact for Researchers in regards to vulnerabilities.

³³⁷ “Complicated or Complex - Knowing the Difference Is Important.”

³³⁸ Lambert, “Disruptive Genius.”

³³⁹ Christensen, *The Innovator’s Dilemma*.

³⁴⁰ Cousins, “Weapons of Mass Disruption.”

³⁴¹ Barker, “Recommendation for Key Management: Part 1: General (Revision 4) DRAFT SP800-57.”

³⁴² Wheeler, “Secure Software Design & Programming - Formal Methods.”

³⁴³ Institute for Defense Analyses, “Open Source Software (OSS/FLOSS) and Security International Workshop on Free/Open Source Software Technologies Riyadh, Saudi Arabia.”



19.3 Units & numbers

Metric prefixes (except for data), **SI units** (except temperature) and the **short number scale** are used.

Examples are listed below for clarification

19.3.1 Short number scale

1,000 = one, with 4 significant figures

$1.000 = 1'000 = 1000 = 1 \cdot 10^3 = 1 \cdot 10^3 = 1$ thousand

$1.000.000 = 1'000'000 = 1 \cdot 10^6 = 1 \cdot 10^6 = 1000 \cdot 1000^1 = 1$ million = 1 mil. = 1 mio. (Danish)

$1.000.000.000 = 1'000'000'000 = 1 \cdot 10^9 = 1 \cdot 10^9 = 1000 \cdot 1000^2 = 1$ **billion** = 1 bil. = 1 mia. (Danish)

$1.000.000.000.000 = 1'000'000'000'000 = 1 \cdot 10^{12} = 1 \cdot 10^{12} = 1000 \cdot 1000^3 = 1$ **trillion** = 1 tri.

19.3.2 Metric prefixes

exa	E	1000^6	10^{18}	1000000000000000000	quintillion
peta	P	1000^5	10^{15}	1000000000000000	quadrillion
tera	T	1000^4	10^{12}	1000000000000	trillion
giga	G	1000^3	10^9	1000000000	billion
mega	M	1000^2	10^6	1000000	million
kilo	k	1000^1	10^3	1000	thousand
hecto	h	$1000^{2/3}$	10^2	100	hundred
deca	da	$1000^{1/3}$	10^1	10	ten
		1000^0	10^0	1	one
deci	d	$1000^{-1/3}$	10^{-1}	0.1	tenth
centi	c	$1000^{-2/3}$	10^{-2}	0.01	hundredth
milli	m	1000^{-1}	10^{-3}	0.001	thousandth

19.3.3 Binary prefixes

	JEDEC		IEC	
1024	K	kilo	Ki	kibi
1024 ²	M	mega	Mi	mebi
1024 ³	G	giga	Gi	gibi

1 Byte B = 8 bit b

Bytes are used when dealing with data storage.

1 bit b = 1/8 Byte B

Bit are used when dealing with data traffic.

19.3.4 SI units

Name	Symbol	Quantity
meter	m	length
kilogram	kg	mass
second	s	time
(kelvin	K	temperature)
(ampere	A	electric current)

19.3.4.1 Derived:

Celsius	°C	temperature	273.15 K
---------	----	-------------	----------



minute	min.	time	60 s
hour	h	time	3600 s (60 minutes)
day	day	time	86400 s (24 hours)
year	year	time	31557600 s (365,25 days)
hertz	Hz	frequency	s ⁻¹
volt	V	voltage	kg·m ² ·s ⁻³ ·A ⁻¹

20 Appendix

20.1 Example Certificate

-ALEXANDER

This is the contents of the modified and then unmodified certificates. Starting with the modified certificate, it is annotated where it differs from the original certificate and the changes are explained. The certificate content is base64 decoded and passed through an ASN.1 DER interpreter in order to visualize the data and any indentation in the data column signifies whether elements are members of a sequence.

20.1.1 Modified certificate overview

-ALEXANDER

When modifying the certificate content fields, either by adding more or by changing their length, all length fields of containing sequences has to be updated as well. Beyond this, the following changes have been made, which are marked in yellow (along with the length fields):

- A padding block is added as a Netscape comment, which features two pieces
 - The block features two pieces. The first part is the string "DONOTTOUCH" repeated until the preceding amount of data in the TBS certificate (including header data) has a length that is completely divisible by 512 bits. The second part consisting of the repeated letter "x", is designated to hold the padding result from the brute force tests and will therefore be overwritten when a successful collision is found.
- The validity period is updates to make the certificate valid another month. This is strictly done to illustrate the point that validity periods by themselves do not offer any security measure. The updated expiration timestamp is still marked as "expired" in order to limit damage, should the private key be leaked
- A new public key is inserted. This is the "payload" of the padding attack, it completely replaces the old public key.
- An extra **SAN** field is added, effectively turning the certificate into a wildcard certificate.

Offset	Length	LenByte	Data
0	1419	3	SEQUENCE :
4	1139	3	SEQUENCE :



8	3	1	CONTEXT SPECIFIC (0) :
10	1	1	INTEGER : 2
13	18	1	INTEGER : 11213F1F0A96160C38E38699E2F747AB7F38
33	13	1	SEQUENCE :
35	9	1	OBJECT IDENTIFIER : sha1withRSAEncryption [1.2.840.113549.1.1.5]
46	0	1	NULL :
48	93	1	SEQUENCE :
50	11	1	SET :
52	9	1	SEQUENCE :
54	3	1	OBJECT IDENTIFIER : countryName [2.5.4.6]
59	2	1	PRINTABLE STRING : 'BE'
63	25	1	SET :
65	23	1	SEQUENCE :
67	3	1	OBJECT IDENTIFIER : organizationName [2.5.4.10]
72	16	1	PRINTABLE STRING : 'GlobalSign nv-sa'
90	51	1	SET :
92	49	1	SEQUENCE :
94	3	1	OBJECT IDENTIFIER : commonName [2.5.4.3]
99	42	1	PRINTABLE STRING : 'GlobalSign Organization Validation CA - G2'
143	30	1	SEQUENCE :
145	13	1	UTC TIME : '130320100505Z'
160	13	1	UTC TIME : '150421100505Z'
175	103	1	SEQUENCE :
177	11	1	SET :
179	9	1	SEQUENCE :
181	3	1	OBJECT IDENTIFIER : countryName [2.5.4.6]
186	2	1	PRINTABLE STRING : 'BE'
190	16	1	SET :
192	14	1	SEQUENCE :
194	3	1	OBJECT IDENTIFIER : stateOrProvinceName [2.5.4.8]
199	7	1	PRINTABLE STRING : 'Belgium'
208	17	1	SET :
210	15	1	SEQUENCE :
212	3	1	OBJECT IDENTIFIER : localityName [2.5.4.7]
217	8	1	PRINTABLE STRING : 'Brussels'
227	28	1	SET :
229	26	1	SEQUENCE :
231	3	1	OBJECT IDENTIFIER : organizationName [2.5.4.10]
236	19	1	PRINTABLE STRING : 'European Commission'
257	21	1	SET :
259	19	1	SEQUENCE :



261	3	1	OBJECT IDENTIFIER : commonName [2.5.4.3]
266	12	1	PRINTABLE STRING : 'ec.europa.eu'
280	290	3	SEQUENCE :
284	13	1	SEQUENCE :
286	9	1	OBJECT IDENTIFIER : rsaEncryption [1.2.840.113549.1.1.1]
297	0	1	NULL :
299	271	3	BIT STRING UnusedBits:0:
304	266	3	SEQUENCE :
308	257	3	INTEGER : 00D4B9410C3DA96517FFF638C2690B465729EDA2E7 0A20BB9B1A953BC3537CC7F71C284637502ABDE828 F162B18A3BAED767891C00825A9F93AD5F76F664DA 63EF42CC94479AE905FC7970EF7BB981CF10F911AC FD936FFEB37FDC95B5F09E40CBE9918C6B9F6D9112 3EB252E517724ADE3FC3E3D9A3DD7C1084ADCE43AD 1069D0F8FEDE4E4D7D5D1479ADA69DA93E094925AA 5DE2443DA64AB9C8D179D6F904C8B15A7C4A699349 53073FA4372A247C113DDA2A2B0AEA2D8D11A2F9A0 3CD5361A37A055623E081D2A36BC5B05EC9449B469 24383202DF77E2B08219BEA74B3B5DDA23FD03B0A2 500CE518C0DA644B5F6CA16DA7B4D7A5B88C1A8D71 4009B45B3B
569	3	1	INTEGER : 65537
574	569	3	CONTEXT SPECIFIC (3) :
578	565	3	SEQUENCE :
582	14	1	SEQUENCE :
584	3	1	OBJECT IDENTIFIER : keyUsage [2.5.29.15]
589	1	1	BOOLEAN : 'FF'
592	4	1	OCTET STRING :
594	2	1	BIT STRING UnusedBits:5 : A0
598	76	1	SEQUENCE :
600	3	1	OBJECT IDENTIFIER : certificatePolicies [2.5.29.32]
605	69	1	OCTET STRING :
607	67	1	SEQUENCE :
609	65	1	SEQUENCE :
611	9	1	OBJECT IDENTIFIER : [1.3.6.1.4.1.4146.1.20]
622	52	1	SEQUENCE :
624	50	1	SEQUENCE :
626	8	1	OBJECT IDENTIFIER : cps [1.3.6.1.5.5.7.2.1]
636	38	1	IA5 STRING : 'https://www.globalsign.com/' 'repository/'
676	36	1	SEQUENCE :
678	3	1	OBJECT IDENTIFIER : subjectAltName [2.5.29.17]
683	29	1	OCTET STRING :



685	27	1	SEQUENCE :
687	12	1	CONTEXT SPECIFIC (2) : 65632E6575726F70612E6575
701	11	1	CONTEXT SPECIFIC (2) : 2A2E6575726F70612E6575
714	9	1	SEQUENCE :
716	3	1	OBJECT IDENTIFIER : basicConstraints [2.5.29.19]
721	2	1	OCTET STRING :
723	0	1	SEQUENCE :
725	29	1	SEQUENCE :
727	3	1	OBJECT IDENTIFIER : extKeyUsage [2.5.29.37]
732	22	1	OCTET STRING :
734	20	1	SEQUENCE :
736	8	1	OBJECT IDENTIFIER : serverAuth [1.3.6.1.5.5.7.3.1]
746	8	1	OBJECT IDENTIFIER : clientAuth [1.3.6.1.5.5.7.3.2]
756	69	1	SEQUENCE :
758	3	1	OBJECT IDENTIFIER : cRLDistributionPoints [2.5.29.31]
763	62	1	OCTET STRING :
765	60	1	SEQUENCE :
767	58	1	SEQUENCE :
769	56	1	CONTEXT SPECIFIC (0) :
771	54	1	CONTEXT SPECIFIC (0) :
773	52	1	CONTEXT SPECIFIC (6) : 'http://crl.globalsign.com/g' 's/gsorganizationvalg2.crl'
827	150	2	SEQUENCE :
830	8	1	OBJECT IDENTIFIER : authorityInfoAccess [1.3.6.1.5.5.7.1.1]
840	137	2	OCTET STRING :
843	134	2	SEQUENCE :
846	71	1	SEQUENCE :
848	8	1	OBJECT IDENTIFIER : caIssuers [1.3.6.1.5.5.7.48.2]
858	59	1	CONTEXT SPECIFIC (6) : 'http://secure.globalsign.com/cace' 'rt/gsorganizationvalg2.crt'
919	59	1	SEQUENCE :
921	8	1	OBJECT IDENTIFIER : ocsf [1.3.6.1.5.5.7.48.1]
931	47	1	CONTEXT SPECIFIC (6) : 'http://ocsp2.globalsign.com/gsoorg' 'anizationvalg2'
980	29	1	SEQUENCE :
982	3	1	OBJECT IDENTIFIER : subjectKeyIdentifier [2.5.29.14]
987	22	1	OCTET STRING :
989	20	1	OCTET STRING : BF852CA8B6B51CED3EFB16BF025110B0907971F3
1011	31	1	SEQUENCE :
1013	3	1	OBJECT IDENTIFIER : authorityKeyIdentifier [2.5.29.35]
1018	24	1	OCTET STRING :



1020	22	1	SEQUENCE :
1022	20	1	CONTEXT SPECIFIC (0) : 5D46B28DC44B741CBBEDF573B63AB7388F75 9E7E
1044	101	1	SEQUENCE :
1046	9	1	OBJECT IDENTIFIER : netscape-comment [2.16.840.1.113730.1.13]
1057	88	1	OCTET STRING :
1059	86	1	IA5 STRING : 'DONOTTOUCHDONOTTOUCHDONOTTOUCHDxxxxxxxx' 'xx' 'xxxxxxxx'
1147	13	1	SEQUENCE :
1149	9	1	OBJECT IDENTIFIER : sha1withRSAEncryption [1.2.840.113549.1.1.5]
1160	0	1	NULL :
1162	257	3	BIT STRING UnusedBits:0: 28ADF91FC5E3C97536A013BE2F0E8B4ED5DE4573B070D39E5A18CF 4E43C048E3926B828830ECF883C4D8C7506F1622CB80BA5AE9F553 F604712C9AF5B21E6491BCF496DDA7462CE7CC7ABFB183A629CB76 2F525EA9E3F14A23ED708454C73409784B4279B465D21B7EEAF2E7 131FAB44237C728C9B0D4607594E4C0425A50FCB18F8A10ECF4F14 3389D96F25DBD6AA611C14D01F2DE525F56F14B926871E9644A71E BE2517764D6F0328F6B72585564A02C55D88DCC92CEB769391E2E1 3E0CF5D0C0A0F428FD99C9B7F027D4C96D37D997B8B9D0FF5429C9 A1A15A5CD54E3050F0360C99B55DF5FB9C24AAF53F7E0EFA403047 0F0189393D97D2F955BC55AE82

20.1.1.1 Original certificate

-ALEXANDER

Offset	Length	LenByte	Data
0	1303	3	SEQUENCE :
4	1023	3	SEQUENCE :
8	3	1	CONTEXT SPECIFIC (0) :
10	1	1	INTEGER : 2
13	18	1	INTEGER : 11213F1F0A96160C38E38699E2F747AB7F38
33	13	1	SEQUENCE :
35	9	1	OBJECT IDENTIFIER : sha1withRSAEncryption [1.2.840.113549.1.1.5]
46	0	1	NULL :
48	93	1	SEQUENCE :
50	11	1	SET :
52	9	1	SEQUENCE :
54	3	1	OBJECT IDENTIFIER : countryName [2.5.4.6]
59	2	1	PRINTABLE STRING : 'BE'
63	25	1	SET :



65	23	1	SEQUENCE :
67	3	1	OBJECT IDENTIFIER : organizationName [2.5.4.10]
72	16	1	PRINTABLE STRING : 'GlobalSign nv-sa'
90	51	1	SET :
92	49	1	SEQUENCE :
94	3	1	OBJECT IDENTIFIER : commonName [2.5.4.3]
99	42	1	PRINTABLE STRING : 'GlobalSign Organization Validation CA - G2'
143	30	1	SEQUENCE :
145	13	1	UTC TIME : '130320100505Z'
160	13	1	UTC TIME : '150321100505Z'
175	103	1	SEQUENCE :
177	11	1	SET :
179	9	1	SEQUENCE :
181	3	1	OBJECT IDENTIFIER : countryName [2.5.4.6]
186	2	1	PRINTABLE STRING : 'BE'
190	16	1	SET :
192	14	1	SEQUENCE :
194	3	1	OBJECT IDENTIFIER : stateOrProvinceName [2.5.4.8]
199	7	1	PRINTABLE STRING : 'Belgium'
208	17	1	SET :
210	15	1	SEQUENCE :
212	3	1	OBJECT IDENTIFIER : localityName [2.5.4.7]
217	8	1	PRINTABLE STRING : 'Brussels'
227	28	1	SET :
229	26	1	SEQUENCE :
231	3	1	OBJECT IDENTIFIER : organizationName [2.5.4.10]
236	19	1	PRINTABLE STRING : 'European Commission'
257	21	1	SET :
259	19	1	SEQUENCE :
261	3	1	OBJECT IDENTIFIER : commonName [2.5.4.3]
266	12	1	PRINTABLE STRING : 'ec.europa.eu'
280	290	3	SEQUENCE :
284	13	1	SEQUENCE :
286	9	1	OBJECT IDENTIFIER : rsaEncryption [1.2.840.113549.1.1.1]
297	0	1	NULL :
299	271	3	BIT STRING UnusedBits:0:
304	266	3	SEQUENCE :
308	257	3	INTEGER : 00EC8ECF7EC9359DDA73D3D2A3D6E7F7010B715011 4831D8713394C34CA385297915503FA03866CBB71D 69F520EACD4736C577EB7D57ECE5CB1416B63089D4



			16A21126572A231EF500DEC2ECF637B443616080CB 70E6D27CC35E22BE234DD47AF4D77FBB5E4633CCAC DAC167987DE2876B930A15471E6386AE62DC15BE00 A5934688DFEA840D91082CC7239A98B8C1E41B6DE0 7EB6974B87EDC1B402A045ED05B151ADF17E712240 071EDED71CFA568FBBEBADE478A343B4FBA37CE412 A076DCB4C86B6835850C40ACD842FCDE0CCB3F35F9 34E47C4DCF0546FED10B995F9AE02FE94ABF2F3093 301E17462A9F26D824068ECD100B28E71E722748F5 C831238033
569	3	1	INTEGER : 65537
574	453	3	CONTEXT SPECIFIC (3) :
578	449	3	SEQUENCE :
582	14	1	SEQUENCE :
584	3	1	OBJECT IDENTIFIER : keyUsage [2.5.29.15]
589	1	1	BOOLEAN : 'FF'
592	4	1	OCTET STRING :
594	2	1	BIT STRING UnusedBits:5 : A0
598	76	1	SEQUENCE :
600	3	1	OBJECT IDENTIFIER : certificatePolicies [2.5.29.32]
605	69	1	OCTET STRING :
607	67	1	SEQUENCE :
609	65	1	SEQUENCE :
611	9	1	OBJECT IDENTIFIER : [1.3.6.1.4.1.4146.1.20]
622	52	1	SEQUENCE :
624	50	1	SEQUENCE :
626	8	1	OBJECT IDENTIFIER : cps [1.3.6.1.5.5.7.2.1]
636	38	1	IA5 STRING : 'https://www.globalsign.com/' 'repository/'
676	23	1	SEQUENCE :
678	3	1	OBJECT IDENTIFIER : subjectAltName [2.5.29.17]
683	16	1	OCTET STRING :
685	14	1	SEQUENCE :
687	12	1	CONTEXT SPECIFIC (2) : 65632E6575726F70612E6575
701	9	1	SEQUENCE :
703	3	1	OBJECT IDENTIFIER : basicConstraints [2.5.29.19]
708	2	1	OCTET STRING :
710	0	1	SEQUENCE :
712	29	1	SEQUENCE :
714	3	1	OBJECT IDENTIFIER : extKeyUsage [2.5.29.37]
719	22	1	OCTET STRING :
721	20	1	SEQUENCE :
723	8	1	OBJECT IDENTIFIER : serverAuth [1.3.6.1.5.5.7.3.1]
733	8	1	OBJECT IDENTIFIER : clientAuth [1.3.6.1.5.5.7.3.2]



743	69	1	SEQUENCE :
745	3	1	OBJECT IDENTIFIER : cRLDistributionPoints [2.5.29.31]
750	62	1	OCTET STRING :
752	60	1	SEQUENCE :
754	58	1	SEQUENCE :
756	56	1	CONTEXT SPECIFIC (0) :
758	54	1	CONTEXT SPECIFIC (0) :
760	52	1	CONTEXT SPECIFIC (6) : 'http://crl.globalsign.com/g' 's/gsorganizationvalg2.crl'
814	150	2	SEQUENCE :
817	8	1	OBJECT IDENTIFIER : authorityInfoAccess [1.3.6.1.5.5.7.1.1]
827	137	2	OCTET STRING :
830	134	2	SEQUENCE :
833	71	1	SEQUENCE :
835	8	1	OBJECT IDENTIFIER : calssuers [1.3.6.1.5.5.7.48.2]
845	59	1	CONTEXT SPECIFIC (6) : 'http://secure.globalsign.com/cace' 'rt/gsorganizationvalg2.crt'
906	59	1	SEQUENCE :
908	8	1	OBJECT IDENTIFIER : ocsp [1.3.6.1.5.5.7.48.1]
918	47	1	CONTEXT SPECIFIC (6) : 'http://ocsp2.globalsign.com/gsoorg' 'anizationvalg2'
967	29	1	SEQUENCE :
969	3	1	OBJECT IDENTIFIER : subjectKeyIdentifier [2.5.29.14]
974	22	1	OCTET STRING :
976	20	1	OCTET STRING : BF852CA8B6B51CED3EFB16BF025110B0907971F3
998	31	1	SEQUENCE :
1000	3	1	OBJECT IDENTIFIER : authorityKeyIdentifier [2.5.29.35]
1005	24	1	OCTET STRING :
1007	22	1	SEQUENCE :
1009	20	1	CONTEXT SPECIFIC (0) : 5D46B28DC44B741CBBEDF573B63AB7388F75 9E7E
1031	13	1	SEQUENCE :
1033	9	1	OBJECT IDENTIFIER : sha1withRSAEncryption [1.2.840.113549.1.1.5]
1044	0	1	NULL :
1046	257	3	BIT STRING UnusedBits:0: 28ADF91FC5E3C97536A013BE2F0E8B4ED5DE4573B070D39E5A18CF 4E43C048E3926B828830ECF883C4D8C7506F1622CB80BA5AE9F553 F604712C9AF5B21E6491BCF496DDA7462CE7CC7ABFB183A629CB76 2F525EA9E3F14A23ED708454C73409784B4279B465D21B7EEAF2E7 131FAB44237C728C9B0D4607594E4C0425A50FCB18F8A10ECF4F14 3389D96F25DBD6AA611C14D01F2DE525F56F14B926871E9644A71E BE2517764D6F0328F6B72585564A02C55D88DCC92CEB769391E2E1



			3E0CF5D0C0A0F428FD99C9B7F027D4C96D37D997B8B9D0FF5429C9 A1A15A5CD54E3050F0360C99B55DF5FB9C24AAF53F7E0EFA403047 0F0189393D97D2F955BC55AE82
--	--	--	--

20.2 Bencoding

-ALEXANDER

In order to ensure a uniform performance across platforms, torrent files are encoded in the Bencode format, based on **UTF-8** strings, with all entries in plaintext. Cohen, “The BitTorrent Protocol Specification.” (p. Bencoding). In order to decode and parse .torrent files, the format must be understood.

Bencoding supports 4 constructs, which mark up the content of the file: Strings, Integers, Lists and Dictionaries, with each type having a unique prefix allowing them to be parsed by a simple stack-based decoder.

Using these four types of data, more complex types can be expressed, as either lists or dictionaries of items, their encoding works as follows:

Strings: <length of string>:<string data>. Here string data can be any string.

- Integers: i<integer>e. Here the integer can be any valid integer between negative and positive infinity. Leading zeroes are not allowed and neither is negative zero.
- List: l<list element 1><list element 2>...<list element n>e. Lists can contain an arbitrary amount of data. Note that there are no element separators since each element clearly mark its own termination.
- Dictionary: d<key 1><value 1><key n><value n>e. Dictionaries perform just like lists, with the added constraint that elements must come in key/value pairs and keys must be of string types.

Note that no intermediate markers are used and whitespace is ignored (except in string literals).

20.3 HPC Platform Deployment

-ALEXANDER

Deploying to the ABACUS 2.0 **HPC**, is done by interfacing with the **SLURM**^{344 345} Batch manager, which manages job submission and execution.

Each user of the **HPC** has an account associated which keeps track of how many node-hours are currently available and a user can consume no more than this. **SLURM** allows a user to submit and manage **HPC** jobs as well as fine tune the amount of resources available to the particular job.

20.3.1 Job scripts

-ALEXANDER

SLURM Job scripts are formatted like standard bash scripts, and the computation time made available is for the entire script.

There are however some specific features made available by **SLURM** to allow finer control of the operation by prefixing a script with **SLURM** directives. These include:

³⁴⁴ “Simple Linux Utility for Resource Management.”

³⁴⁵ “Slurm Job Scheduler.”



- Account management:
 - The directive “#SBATCH – account <account name>” sets the account node hour pool the user draws resources from. This allows user to participate in multiple projects while still drawing from correct pools.
- Node scalability:
 - The directive “#SBATCH –nodes <number>” or “#SBATCH –nodes <min>-<max>” allows for control over how many nodes a project will at minimum require before running as well as the maximum the project can utilize. If only a single number is set, then the script will not execute before the manager can schedule that amount of simultaneous nodes at once.
- Timing constrains:
 - The directive “#SBATCH –time <hh:mm:ss>” indicates the maximum amount of execution time each node will before being forcefully terminated. For jobs that run until cancelled this is the approximate time the job will run, but it will not be an exact number since the scheduler spends some time updating gathering input/output before moving a job in/out of the schedule queue.
- Output redirection:
 - Since slurm can execute the same script on multiple nodes, it may make sense to redirect the output to a dedicated log instead of the default output stream. This is done with the directive “#SBATCH --output=<name>”. Here the name can feature special markers such as the job id, in order to easier differentiate submitted jobs.
- Execution variables:
 - Such as the submission directory, the active nodelist, a unique node id or the likes, allowing jobs to be aware of and utilize the fact that they are running on a multimode cluster.

20.3.2 ABACUS Scripts

-ALEXANDER

The scripts used to interface with ABACUS are as follows:

- “slurmscript_diag.sh” which is a script executing **CUDA** diagnostics instructions. These have to be deployed to the **GPU** nodes instead of the login and development frontend, since the frontend does not feature any test **GPU** environment, and any diagnostics information would be unavailable. This is done through the “nvidia-smi” driver interface, which allows access to a complete diagnostics of hardware parameters such as power and temperature readings as well as memory and SMX usage. While not critical for the project this is excellent information for troubleshooting and performance evaluation purposes and it is therefore kept as a part of the deployment process.
- “slurmscript_flexible”, “slurmscript_full”, “slurmscriptopt_flexible” and “slurmscriptopt_full”. These four scripts launch the non-optimized and optimized version of the **CUDA** application in either full or flexible configuration. Full and flexible refer to the **SLURM** node configuration used. When using the full configuration, the script requests access to all 72 nodes before launching and the queue system will wait until all nodes are available before giving the application time to execute. This can potentially stall the execution for a long time. The flexible configuration will execute when between 1 and 72 nodes are available, and is therefore expected to execute sooner, but with a reduced amount of nodes, the computational power will be proportionally reduced.
- The “test.sh” script compiles the **CUDA** executables and launches all the **SLURM** scripts. It is the main script used for deployment and testing, hence the name.



20.4 Shamir Secret Sharing Toolkit Readme

-LARS

Thank you for helping us out.

To use / Install:

Start Kali linux live CD

Move files to any folder you would like

chmod (eg. chmod 777 verifySHA.sh)

run the bash script (eg. ./verifySHA.sh)

>If you are combining or verifying,

>read what folder the files should be in below

>(this can be changed in the first few lines of the script)

This USB pen contains 3 bash scripts:

Combine.sh

>Can combine Shamir secrets (shards) to the original file

>Default folder: ~/RSA/SHARD_OUTPUT

verifySHA.sh

>Verifies an RSA signature (ec.crt.sig)

>Default folder: ~/RSA

All_in_one.sh

>Replicating the procedure used to generate the private key and the Shamir secret (shard) on this USB pen.

>Please contact us if you find any security errors in the procedure.

>Default folder: N/A

>it will install into ~/RSA, but the All_in_one.sh can be executed from anywhere on the system.

Furthermore this USB pen contains:

public.pem

>The public key corresponding to the private RSA key

keys.txt

>Yours and our public openPGP keys

libgfshare-2.0.0.tar.bz2

>Installation files for Shamir secret code, by Daniel Silverstone

><https://git.gitano.org.uk/libgfshare.git/snapshot/libgfshare-2.0.0.tar.bz2>

automake-1.15.tar.gz, autoconf-2.69.tar.gz, libtool-2.4.6.tar.gz

>Dependancies for the libgfshare Shamir tool, that are not part of Kali linux

>so this script can run on an air gapped computer

><https://ftp.gnu.org/gnu/automake/automake-1.15.tar.gz>

><https://ftp.gnu.org/gnu/autoconf/autoconf-2.69.tar.gz>

><http://ftpmirror.gnu.org/libtool/libtool-2.4.6.tar.gz>

ec.crt



>European Commission certificate, as validated by the Danish Trusted Service List Version 4, sequence 9 of 2015/03/20T07>53>10Z

ec.crt.sig

>Signature of European Commission certificate with our private key (to verify existence and ownership of private key by signing a "Nothing up my sleeve" value

SHA512_hash_ec.crt

>SHA512 of European Commission certificate, to ease verification the signature.

20.5 ERA letters

20.5.1 Letter 1, December 2nd 12:02

Subject: FW: Information Request Form - Nielsen

Date: Wed, 2 Dec 2015 11:02:19 +0000

From: Communication <Communication@era.europa.eu>

To: s042903@student.dtu.dk <s042903@student.dtu.dk>

Dear Mr. Nielsen,

First of all we would like to thank you for your interest in the ERA activities.

ERA has no responsibility concerning IT security policies, our scope of work is related only to railway safety and interoperability. The European Agency dealing with security is ENISA (European Union Agency for Network and Information Security). Please contact them at <https://www.enisa.europa.eu/>

Concerning ERTMS, the IT security is based on the exchange of keys between interoperability constituents and the protocols defined for Euroradio. All the specifications can be downloaded at: <http://www.era.europa.eu/Core-Activities/ERTMS/Pages/Set-of-specifications-2.aspx>

More in details, please be aware that in the particular case that you refer to (STM interface), as it is an interface between the EVC and the STM module and both are located at the train, at least you would need:

- to get physical access to the cab train,
- to be able to power up a train,
- to introduce the correct parameters for a train mission,
- to hack the interface,
- to provide correct signalling information.

In any case, the security level can always be increased by national governments and/or railway managers, if they consider to do so, by implementing further security protocols.

Best regards,

EUROPEAN RAILWAY AGENCY
Corporate Management and Evaluation



Communication Office
120 rue Marc Lefrancq
BP20392
FR-59307 VALENCIENNES Cedex
tel: 00.33.(0)3.27.09.65.00
www.era.europa.eu

If you want to subscribe to the European Railway Agency's Flash News follow this link.

-----Original Message-----

From: Lars Nielsen [mailto:s042903@student.dtu.dk]
Sent: Thursday, November 12, 2015 10:17 AM
To: SENECHAL H el ene (ERA)
Cc: Alexander Adelholm Brandbyge
Subject: Re: FW: urgent: Information Request Form - Nielsen

Hi H el ene Senechal

short:

The ERTMS standard conflicts with the NIST security advice.

More precise:

STM FFFIS Safe Link Layer section 5.2.3.4 specifying that the authentication token is only 32 bits, with an unknown/unspecified algorithm (section 5.2.3, 5.1.4) [1]

While NIST SP800-57 recommends -at least- 80 bits (in legacy mode) and 112+ bits. [2](page 8) [3] (Table 4, page 67)

Furthermore describing that truncated digests need to have an improved hashing algorithm. [4](page 9-10)

We hope this is specific enough?

The questions are:

- A) What standards / recommendations does ERA recommend in regards to IT security in the railway sector?
- B) How does ERA think broken hashing algorithms affect the railway sector? (as in risk analysis)
- C) What is the plan for specifying and upgrading IT security measures in ERTMS (given that railway systems usually run for decades)
- D) What is the reasoning behind thinking that NIST recommendations are too strict, when they are based on security research papers proving low entropy (few bits) can be broken. [5]

Kind Regards

Alexander Adelholm Brandbyge <s143358@student.dtu.dk>

&

Lars Emb oll Nielsen <s042903@student.dtu.dk>

[1]STM FFFIS Safe Link Layer (

<http://www.era.europa.eu/Document-Register/Pages/STM-FFFIS-Safe-link-Layer.aspx>

)



- [2] SP 800-131 A Transitions: Recommendation for Transitioning the Use of Cryptographic Algorithms and Key Lengths (
<http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-131Ar1.pdf>
)
- [3] SP 800-57 Part 1 Recommendation for Key Management: Part 1: General (
http://csrc.nist.gov/publications/nistpubs/800-57/sp800-57_part1_rev3_general.pdf
)
- [4] NIST SP 800-107 Recommendation for Applications Using Approved Hash Algorithms (
<http://csrc.nist.gov/publications/nistpubs/800-107-rev1/sp800-107-rev1.pdf>)
- [5] Marc Stevens, Jacob Appelbaum - MD5 considered harmful today (
<https://www.win.tue.nl/hashclash/rogue-ca/>)
- [6] George Argyros - Aggelos Kiayias -PRNG: Pwning Random Number Generators(
https://media.blackhat.com/bh-us-12/Briefings/Argyros/BH_US_12_Argyros_PRNG_WP.pdf
)

On 10/11-2015 15:28, SENECHAL H el ene (ERA) wrote:

```
> * *
>
> Dear Mr Nielsen
>
>
>
> Thank you for your interest in our Agency's work.
>
> Currently we are processing your request, but unfortunately it seems it
> may take a bit longer than expected.
>
> Could you please be more precise in order to help our Project officer to
> find the right answer?
>
> Thank you in advance for your patience and for your cooperation.
> Hi ERA, we are writing our master thesis at the Technical University of
> Denmark (www.DTU.dk <http://www.DTU.dk>) on risk assessment of hashing
> algorithms with Christian Damsgaard[1] as our supervisor. One of our
> research goals is to cover the railway sector, to estimate the impact a
> broken hashing algorithm would have on that domain. We have had a hard
> time getting contacts within this sector, could you be helpful with: a
> list of possible contacts? relevant design documents? We already have
> obtained 1000 CPU hours on HPC (High Performance Computers) in both
> Iceland and Denmark to run our evaluation of the SHA-1 hashing
> algorithm, with the improved code we have designed. So we have the
> probability part of the risk assessment set. but we are missing data on
> impact / consequence. Kind Regards -Alexander Brandbyge & Lars Nielsen
> [1] Christian Damsgaard Jensen (
> http://www.dtu.dk/Service/Telefonbog/Person?id=13409 )
> -----
>
>
```




- > This message is intended for the use of the addressee only and may
- > contain information that is privileged and/or confidential information.
- > If you are not the intended recipient, you are informed that any
- > dissemination or other use of this message is strictly prohibited. If
- > you have received this message in error, please inform the European
- > Railway Agency immediately by returning it and then delete the material.
- >
- > The European Railway Agency endeavours to keep its network free of
- > viruses; however you are strongly advised to check this e-mail and any
- > attachments for viruses. The European Railway Agency accepts no
- > responsibility with regard to any computer virus transferred by way of
- > this e-mail.

20.5.2 Letter 2, December 3rd 11:49

Subject: FW: Information Request Form - Nielsen
Date: Thu, 3 Dec 2015 10:49:48 +0000
From: Communication <Communication@era.europa.eu>
To: LarsNielsen@RailwayHacker.com <LarsNielsen@RailwayHacker.com>

Dear Mr. Nielsen,

First of all, one clarification. For your scenario "remotely executed attack during regular operation that could eg.

increase the allowed speed, leading to a derailment at a switch/turnout or curve", please consider two issues

1) the ERTMS is not an ATO system i.e. it is a protection system with a driver presence, I mean it is the driver who is driving not the ERTMS system. So, it looks that you would need some cooperation from the driver who needs route knowledge and speed tables to be allowed to drive.

2) Your "fake allowed speed" should come either from and RBC or a balise, so you should know the RBC and balise identifiers and get access to railway installations again.

Please bear in mind that if needed I could even change the keys every time I communicate, so that if you sniff the info it will not be usable for the next communication.

Our specifications does not mention when each key can be changed, it provides the mean to change it. It is up to each administration to do decide when, how often, ...

You could argue that the machine providing the keys can be hacked, of course yes as any IT system, but these machines are normally certified for security and this is beyond the ERTMS and ERA scope of work.

Concerning your last question, please see Subset-037, Subset-038 and Subset-92

Best Regards

-----Original Message-----

From: Lars Nielsen [mailto:LarsNielsen@RailwayHacker.com]
Sent: Wednesday, December 02, 2015 3:11 PM
To: Communication
Subject: Re: FW: Information Request Form - Nielsen



Thank you for a good answer,
I will note down that your risk assessment focus on physical barriers to insure security of the communication.

It is then imperative that vendors implementing this are aware of that and will never use unshielded cables or wireless transmissions for this.

I must admit that I had misunderstood it to be a wireless (EN 50159 Category 3) connection[1]
I apologize for my mistake.
I am glad to see that you have a mitigation in place.

The attack vector I envisioned was a remotely executed attack during regular operation that could eg. increase the allowed speed, leading to a derailment at a switch/turnout or curve.
That is if the security relied on the Safe Link Layer the 4 byte authentication message.

I will direct my questions and future inquiries to ENISA.
Thank you very much for an answer with a good level of details yet very fast response time for a question of technical nature.

I have one final request if possible:
We are aware of where we can download all the specifications, there are quite a lot though, so if you could be more precise on individual specifications regarding IT security it would be appreciated.

To my understanding, and by going through many of those standards (all that seemed to concern communication) the Safe Link Layer was the one dealing with authentication, apart from the GSM-R communication that to my knowledge shares the same attack surface as GSM.

Kind Regards
-Lars Embøll Nielsen

[1] "Category 3 consists of systems which are not under the control of the designer, and where unauthorized access has to be considered" from EN 50159 Railway applications - Communication, signalling and processing systems - Safety-related communication in transmission systems

On 2/12-2015 12:02, Communication wrote:

> Dear Mr. Nielsen,

>

> First of all we would like to thank you for your interest in the ERA activities.

>

> ERA has no responsibility concerning IT security policies, our scope of work is related only to railway safety and interoperability. The European Agency dealing with security is ENISA (European Union Agency for Network and Information Security). Please contact them at <https://www.enisa.europa.eu/>

>



> Concerning ERTMS, the IT security is based on the exchange of keys between interoperability constituents and the protocols defined for Euroradio. All the specifications can be downloaded at:

> <http://www.era.europa.eu/Core-Activities/ERTMS/Pages/Set-of-specifications-2.aspx>

>

> More in details, please be aware that in the particular case that you refer to (STM interface), as it is an interface between the EVC and the STM module and both are located at the train, at least you would need:

> -to get physical access to the cab train,

> -to be able to power up a train,

> -to introduce the correct parameters for a train mission,

> -to hack the interface,

> -to provide correct signalling information.

>

> In any case, the security level can always be increased by national governments and/or railway managers, if they consider to do so, by implementing further security protocols.

>

> Best regards,

>

>

> EUROPEAN RAILWAY AGENCY

> Corporate Management and Evaluation

> Communication Office

> 120 rue Marc Lefrancq

> BP20392

> FR-59307 VALENCIENNES Cedex

> tel: 00.33.(0)3.27.09.65.00

> www.era.europa.eu

> If you want to subscribe to the European Railway Agency's Flash News follow this link.

>

>

>

> -----Original Message-----

> From: Lars Nielsen [mailto:s042903@student.dtu.dk]

> Sent: Thursday, November 12, 2015 10:17 AM

> To: SENECHAL H el ene (ERA)

> Cc: Alexander Adelholm Brandbyge

> Subject: Re: FW: urgent: Information Request Form - Nielsen

>

> Hi H el ene Senechal

> short:

> The ERTMS standard conflicts with the NIST security advice.

>

> More precise:

> STM FFFIS Safe Link Layer section 5.2.3.4 specifying that the

> authentication token is only 32 bits, with an unknown/unspecified

> algorithm (section 5.2.3, 5.1.4) [1]

> While NIST SP800-57 recommends -at least- 80 bits (in legacy mode) and

> 112+ bits. [2](page 8) [3] (Table 4, page 67)

> Furthermore describing that truncated digests need to have an improved

> hashing algorithm. [4](page 9-10)

>

> We hope this is specific enough?



>

> The questions are:

> A) What standards / recommendations does ERA recommend in regards to IT security in the railway sector?

> B) How does ERA think broken hashing algorithms affect the railway sector? (as in risk analysis)

> C) What is the plan for specifying and upgrading IT security measures in ERTMS (given that railway systems usually run for decades)

> D) What is the reasoning behind thinking that NIST recommendations are too strict, when they are based on security research papers proving low entropy (few bits) can be broken. [5]

>

> Kind Regards

> Alexander Adelholm Brandbyge <s143358@student.dtu.dk>

> &

> Lars Embøll Nielsen <s042903@student.dtu.dk>

>

> [1]STM FFFIS Safe Link Layer (
> <http://www.era.europa.eu/Document-Register/Pages/STM-FFFIS-Safe-link-Layer.aspx>
>)

> [2] SP 800-131 A Transitions: Recommendation for Transitioning the Use of Cryptographic Algorithms and Key Lengths (
> <http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-131Ar1.pdf>
>)

> [3] SP 800-57 Part 1 Recommendation for Key Management: Part 1: General
> (
> http://csrc.nist.gov/publications/nistpubs/800-57/sp800-57_part1_rev3_general.pdf
>)

> [4] NIST SP 800-107 Recommendation for Applications Using Approved Hash Algorithms (
> <http://csrc.nist.gov/publications/nistpubs/800-107-rev1/sp800-107-rev1.pdf>)

> [5] Marc Stevens, Jacob Appelbaum - MD5 considered harmful today (
> <https://www.win.tue.nl/hashclash/rogue-ca/>)

> [6] George Argyros - Aggelos Kiayias -PRNG: Pwning Random Number Generators(
> https://media.blackhat.com/bh-us-12/Briefings/Argyros/BH_US_12_Argyros_PRNG_WP.pdf
>)

>

> On 10/11-2015 15:28, SENECHAL H el ene (ERA) wrote:

>> * *

>>

>> Dear Mr Nielsen

>>

>>

>>

>> Thank you for your interest in our Agency's work.

>>

>> Currently we are processing your request, but unfortunately it seems it may take a bit longer than expected.

>>

>> Could you please be more precise in order to help our Project officer to



>> find the right answer?
>>
>> Thank you in advance for your patience and for your cooperation.
>>
>>
>>
>> Best regards
>>
>>
>>
>> *logo_email.png*
>>
>> *EUROPEAN RAILWAY AGENCY*
>>
>> Corporate Management and Evaluation
>>
>> Communication Office
>>
>> 120 rue Marc Lefrancq
>>
>> BP20392
>>
>> FR-59307 VALENCIENNES Cedex
>>
>> tel: 00.33.(0)3.27.09.65.00
>>
>> www.era.europa.eu <<http://www.era.europa.eu/>>
>>
>> If you want to subscribe to the European Railway Agency's Flash News
>> followthislink
>> <<http://www.era.europa.eu/Communication/Newsletter/Pages/home.aspx>>.
>> Hi ERA, we are writing our master thesis at the Technical University of
>> Denmark (www.DTU.dk <<http://www.DTU.dk>>) on risk assessment of hashing
>> algorithms with Christian Damsgaard[1] as our supervisor. One of our
>> research goals is to cover the railway sector, to estimate the impact a
>> broken hashing algorithm would have on that domain. We have had a hard
>> time getting contacts within this sector, could you be helpful with: a
>> list of possible contacts? relevant design documents? We already have
>> obtained 1000 CPU hours on HPC (High Performance Computers) in both
>> Iceland and Denmark to run our evaluation of the SHA-1 hashing
>> algorithm, with the improved code we have designed. So we have the
>> probability part of the risk assessment set. but we are missing data on
>> impact / consequence. Kind Regards -Alexander Brandbyge & Lars Nielsen
>> [1] Christian DamsgaardJensen (
>> <http://www.dtu.dk/Service/Telefonbog/Person?id=13409>)
>> -----
>>
>>
>> This message is intended for the use of the addressee only and may
>> contain information that is privileged and/or confidential information.
>> If you are not the intended recipient, you are informed that any



>> dissemination or other use of this message is strictly prohibited. If
>> you have received this message in error, please inform the European
>> Railway Agency immediately by returning it and then delete the material.
>>
>> The European Railway Agency endeavours to keep its network free of
>> viruses; however you are strongly advised to check this e-mail and any
>> attachments for viruses. The European Railway Agency accepts no
>> responsibility with regard to any computer virus transferred by way of
>> this e-mail.

PAGE 134 OF 134
SIGNATURES
January 11th 2016

ALEXANDER ADELHOLM BRANDBYGE

LARS EMBØLL NIELSEN