University of Tartu

School of Economics and Business Administration

Rauno Oja

CALCULATING THE RETURN ON SECURITY INVESTMENT OF RECODING X-ROAD AND ESTONIAN ELECTRONIC IDENTITY SOFTWARE INTO BLOCKCHAIN

Master Thesis

Supervisor: Researcher Oliver Lukason

Tartu 2016

Recommended for defense ……

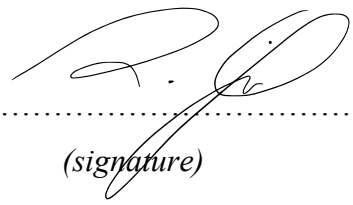
Oliver Lukason

Accepted for defense “ “..... 2016.

…...… Head of chair ……

Urmas Varblane

I have written the Master Thesis myself, independently. All of the other authors’ texts, main viewpoints and all data from other resources have been referred to.



……..

(signature)

……..

(date)

Rauno Oja

Abstract

Online security is essential to any governmental e-service, and more so in cases when electronic identification is used to cast votes in elections or perform financial transactions. Recently, one of the most promising and disruptive innovations in improving the governance of data has been the introduction of blockchain (the technology underpinning Bitcoin). This thesis considers the costs of recoding X-road and Estonian Electronic Identity Software source codes into blockchain using the COCOMO II software cost estimation model. These costs are then used to calculate the return on security investment (ROSI). The results indicate that the potential security risks would have significantly higher costs compared to the to-be incurring costs of implementing blockchain recoding, meaning the ROSI would very strongly justify making the transition. This thesis aims to provide a tangible case for calculating the transfer costs to blockchain technology and evaluates the potential ROSI.

Keywords: *blockchain, ROSI, e-government, x-road, electronic identification, COCOMO.*

1. Introduction

The objective of this thesis is to determine the costs of recoding the source codes of both X-road, the data communication backbone, and eID, the electronic identification software, into blockchain and weigh these development costs against potential security breach costs, expressed in return on security investment (ROSI). The COCOMO II software development cost estimation tool will be used for the former and an existing method of estimating the return on investment in security software improvements for the latter.

Firstly, the thesis provides a general introduction into blockchain technology and into the two products of Estonian e-government to be evaluated for transfer. The reasoning and novelty behind such transmission are considered and discussed.

Secondly, a review of existing literature on transferring to blockchain as well as existing software cost models and various methods for calculating ROSI will be provided. The COCOMO II model is explained in further depth.

The Data and Methodology section of the thesis will describe the assumptions that are used to evaluate the scope of the effort and to calculate the potential costs, but also to estimate the potential financial loss from security breach. It will also consider three scenarios in order to contemplate any cost cutting efforts for the transfer. ROSI is calculated and justifications are provided for the results. The ROSI is also compared to all data available from similar projects or approaches.

Lastly, the author will draw conclusions on the results and provide input on the justifiability of the transfer to blockchain, and also any considerations to be pondered that may affect the results when implementing this transition in real life.

Although software development cost estimation is a fairly mature topic, it is still extremely difficult to estimate accurately due to the quickly developing methods and strain of predicting the actual efforts to create the software. Similarly, decentralized database structure has existed for a while now, but only with the resurgence of blockchain based cryptocurrencies, has it become a more thoroughly researched topic. Return on investment in IT-solutions and costs of risk from potential security break-ins are vastly understudied and often inaccurate. Because this thesis considers blockchain technology

as a solution, it furthermore fills the existing gap in lack of adequate research in assessing transfer costs to blockchain technology for security purposes.

Bitcoin (Nakamoto 2008), the much talked about cryptocurrency (Böhme et al. 2015: 213-238, Popper 2015), has both been cheered as a revolutionary technology in finance and banking (Vigna & Casey 2015) and even called the 'next internet' (Brodbeck 2015), but also dismissed as hoax (Matthews 2014) and a house of cards about to be blown over (Bloomberg 2016). While speculation over the price of Bitcoin and arguments about its value as a currency continue, many still agree blockchain, the underlying technology behind Bitcoin, is the next big thing (Tapscott 2015, The Economist 2015, Swan 2015: vii). Blockchain could also be used as an underlying technology in forming smart contracts, which essentially allows both identification of the participant and agreement on a set of terms within the contract (Kosba et al. 2015). However, it's not just the world of finance that blockchain is reshaping, but also the way third party verification is used as a whole in banking, real estate, health care and data management (The Economist 2015, Lavinskaya 2016, Mougayar 2014, Swan 2015: 40,76, UK Office for Science 2015: 14).

Blockchain is a database, acting as a distributed ledger (Swanson 2015) with real time, nearly tamper-proof processing capability, which in addition is also increasingly low-cost (UK Office for Science 2015: 6). It has immense potential to revolutionize *modus operandi* of governments to provide their services (Naughton 2016). The reason for that is blockchain's ability to remove a single trusted third party from the verification process (Buterin 2015, Plassaras 2013, Back et al. 2014, Bissias et al. 2014). Many governments are already closely looking into policy-making for blockchain technologies and even encouraging people to use it (Baraniuk 2016).

The first electronic identification system was introduced as early as the 1960's (Enikeieff et al. 1965) and further improvements were patented in the late 1980's, using simple pre-established secret keys or personal identification codes (PIN's), and was mainly used for approving fund transfers in ATM's (Stein 1989). Although many improvements have since been introduced to the ID card holder's verification process itself (e.g. fingerprints (Lane 1997) and biometrics (Berson & Zemlok 1995)), bio-information scanning technology is too expensive to be implemented into every household and also faces security concerns of its own.

Target asset security is only as secure as the communication channel between the identity holder and the identity provider (Neumann 2013: 21-32). Carter and Bélanger (2005: 5-25) claim that trustworthiness, alongside ease of use and compatibility, is a significant factor for a citizen to determine their use of governmental e-services and the success or acceptance of such services amongst the population. West (2004) also argues that privacy and security are clearly reflected as area of concern in surveys conducted in US about their use of e-government services.

Estonia introduced X-road, the secure data transport backbone and various distributed software systems to create a common system for e-Government (Kalja et al. 2005). Ever since, the country has been the leading advocate of e-services and their digital signature and the identification system, which helps citizens to vote online, access digital health records and monitor any data inquires by the government, and has received a a great deal of praise in the mainstream media (Scott 2014, Mansel 2013). The e-state project has also caught on and countries such as Japan, Ukraine, Finland and Georgia have started collaboration to introduce similar digital revolutions to their citizens as well (ERR 2013, BNS 2015, e-Estonia.com 2015). Despite all the kind words, there are still recognizable issues with the system to be tackled (Deane 2003, Kitsing 2011).

2. Literature Review

This section covers the three main components of the thesis – 1) coding to Blockchain, 2) COCOMO II cost model and 3) ROSI and overview of existing research and publications.

2.1. Transition to Blockchain

Blockchain is still very much an emerging technology and not much scientific literature is available on its properties, nor is there a wide range of case studies available for review. Consequently, it is difficult also to find any harmonized guidelines for implementing blockchain for diverse applications, and more so on a larger scale. There have however been a couple of experimental instructions (Czepluch et al. 2015, Spanos 2015) for attempts to code blockchain applications, but also much development of the method has been done by Ethereum, a decentralized platform for running smart contracts or applications run on a custom built blockchain.

Vitalik Buterin (2013: 14-21), the founder of Ethereum, explains the code execution as: "an infinite loop that consists of repeatedly carrying out the operation at the current program counter (which begins at zero) and then incrementing the program counter by one, until the end of the code is reached or an error or STOP or RETURN instruction is detected." Essentially though, the coding is still a series of instructions that in the case of blockchain are based around a logic of zero-knowledge proof (Buterin 2016). Therefore, one can examine the existing source code and assume certain additional efforts by programmers and developers to recode the existing code into blockchain.

Use of the Ethereum mechanism is also encouraged by Czepluch et al. (2015: 3) on the basis of the transparency of the method, which allows accurate verification of the predicted transactions. Still, a notable concern is the scalability of the method, as current estimations and suggestions for large scale projects are only theories and are still untested, and the blockchain may become too large for an average user to carry a node on to the network (Czepluch et al. 2015: 60). Trent et al. (2016: 50-56) address this in their research and conclude that blockchain is capable of running a large scale decentralized database and offer a solution of their own in the form of BigchainDB as a building block within blockchain applications to run the database.

2.2. COCOMO II

Computer programming cost estimation dates back to as early as the mid-60s (Nelson 1966), which led to some first generation models in the late 60's-early 70's (Boehm et al. 2000). The late 1970's produced improved methods with SLIM, Checkpoint, PRICE-S, SEER and COCOMO models being produced (Boehm et al. 2000). Software development cost estimation has been a very difficult task to perform accurately, mainly due to the inability to accurately anticipate skill sets, unexpected additional tasks, change in set tasks connected to the fast pace of development in technology and pressures from managements to increase/decrease the estimates (Lederer & Prasad 1995). Regression-based estimation techniques, such as COCOMO modelling, seem to dominate by far in use for all assessments (Jørgensen & Shepperd 2007), however a composite method introduced with the COCOMO II model in 1995 is more appropriate and precise for projects with top-level design in place (Boehm et al. 2000). The Post-Architecture model of the COCOMO II has also been calibrated with 161 projects from both the private and public sector, including also non-profit organizations (Boehm et al. 2000), making it most appropriate for the purpose of this thesis.

COCOMO II (or COCOMO 2.0) uses source instructions (source code) and sets of five Scale Factors (SF) and 17 Cost Drivers or Effort Multipliers (EM) to determine the overall development costs of a project (USC 2000, Boehm et al. 1995) (Appendix1). This is best suited for software which already has developed its life-cycle architecture, validated its' purpose and established a framework (Boehm et al. 1995). Because the X-road and Estonian Electronic Identity Software have already been up and running for some time and considered for mere alteration for purposes of recoding, the COCOMO 2.0 model fits well for the motivation of the thesis.

Software cost drivers are also factors that capture the specifics of the software, which influence the effort that goes into creating the program and drive the costs (expressed in Person Months) (USC 2000). As with Scale Factors, each driver is given a rating from very low to extra high (with some exceptions - see Appendix 1). Special consideration should be given to the schedule factor, as the end effort is expressed in person months (PM) and therefore the schedule estimates and precedents have major importance in the result.

Project size is one the most important inputs in COCOMO II models and has a special cost driver, Exponent E (USC 2000). It is an aggregation of all five scale factors (Appendix 1).

2.3. ROSI

Several models have been proposed for calculation of returns on investment in IT-security (Cavusoglu et al. 2004, Al-Humaigani & Dunn 2003, Tsiakis & Pekos 2008, Sonnenreich 2006). The majority of them concentrate on the potential risk and loss occurred, compared to the investment to be made towards higher security. Often, security considerations are a 'what if' scenario and executive decision makers concentrate too much on the bottom-line scenarios (Sonnenreich 2006). From a managerial point of view, each investment (including security) also has to make sense financially, although often the intruders are not only financially motivated (Cavusoglu et al. 2004).

Kramer et al. (2015: 186) indicate that finding errors and resolving them in the early stages of the development will significantly decrease the end costs of the project. Therefore, it is worthwhile to invest in an error-checking labor force in the initial phase. Similarly, the ROSI models described by Tsalis et al. (2013: 133-134) factor in the efforts of mitigation in case of a security risk. Their case study offers a good example of calculating ROSI as it deals with transferring data to cloud service, considers the development costs and focuses on security controls (Tsalis et al. 2013: 136).

Perhaps the best guidelines are provided by the European Network and Information Security Agency (ENISA 2012), who provide an introductory overview to the world of calculating the ROSI. The authors do emphasize that calculations are not about financial gain, but rather about prevention of loss (ENISA 2012: 2). The report also indicates the potential drawbacks of estimating ROSI, namely that it is a result of many approximations, and advise to refer to historic data on specific cases, even though they can be difficult to come by as companies are reluctant to share that information (ENISA 2012: 7). There are models, such as White Paper on Measuring the Return on IT Security Investments from Intel (Rosenquist 2007: 5), which do provide a scalable model, but also rely on the method of using large amounts of historic data from previous incidents, which couldn't be applied across the board for all cases.

Return on investment in security differs from regular ROI calculations. Rather than counting the benefits, the avoided cost in data breach is compared to the cost of investment (Booz et al. 2014). In order to calculate the ROSI, a potential threat should also be assumed. For the purposes of this thesis, ROSI will be calculated assuming a malicious outside cyber attack on the system, that is within the frameworks of similar attacks on large corporations or governmental institutions.

In order to calculate the ROSI or Return on Security Investment, one should estimate the costs of countermeasures (CC), i.e., the cost of solution (SC) and evaluate the Annual Loss Expectancy (ALE) (Sonnenreich 2006, ENISA 2012):

$$ROSI = \frac{(ALE \times \%Risk\ Mitigated) - Solution\ Cost}{Solution\ Cost}$$

Percentage of Risk Mitigation (%Risk Mitigated) reflects the proportion of attacks the solution is intended to address.

A basic formula to capture the costs of data breach is:

ALE (Annual Loss Expectancy) = SLE (Single Loss Expectancy) ×ARO (Annualized Rate of Occurrence) (Stuntz 2014, Krause & Tipton 1993).

Single Loss Expectancy (SLE) expresses the monetary impact of a single event or threat (Krause & Tipton 1993) and is defined as:

SLE = ASSET VALUE × EXPOSURE FACTOR

Whereas, exposure factor is a measure of extent of the loss expressed in percentage.

3. Data and Methodology

In order to estimate the total costs of the development at hand, the following steps should be made:

1) Identify the total volume of the existing source code.
2) Identify and assign scale factors and effort multipliers based on the specifications given in the COCOMO II guidelines and provided in Appendix 1.
3) Calculate the amount of source code to be recoded in order to work on blockchain.
4) Calculate the total effort that would go into the recoding.
5) Calculate how much time it would take to put in this effort.
6) Calculate how much it would cost to put in this effort.

Additionally, several approaches will be identified in this section which could reduce the cost for development and recommendations for prior actions will be given that could produce the most cost effective results.

3.1. Existing source code

For the purposes of this thesis, a total of twenty repositories available for public access (Population Register Center 2016, Riigi Infosüsteemi Amet (RIA) 2016) with the source code for the X-road and Electronic Identity Software, were used. X-Road is the software that facilitates information transfer between parties and the Electronic Identity Software manages the requests and adds signatures (validity) to documents once electronically signed. The software also includes a digidoc program, which is downloaded onto the users' PC to sign documents. The following depositories were used:

1) \browser-token-signing-master
2) \esteid-pkcs11-master
3) \libdigidoc-master
4) \ndigidoc-master
5) \chrome-token-signing-master
6) \esteid-tokend-master
7) \google-breakpad-r1403
8) \libdigidocpp-master
9) \minidriver-master
10) \qdigidoc-master
11) \qesteidutil-master
12) \qt-common-master
13) \smartcardpp-master
14) \updater-master
15) \digidoc4j-master
16) \dss-hwcrypto-demo-master
17) \jdigidoc-master
18) \pdf-validator-master
19) \sd-dss-master
20) \xroad-public-maste

These folders were then analyzed for the total count of SLOCs (Source Lines of Code) with the LocMetrics counting tool. For the purposes of this thesis, the total amount of Executable Logical (SLOC-L) code lines will be used for calculations, as this indicator best reflects the actual content to be revised (removing blank lines, comment lines and singular symbol lines) (Nguyen et al. 2007). The total number of Executable Logical SLOC's to be used is 222623 (Figure 1: LocMetrics Count of Source Lines of Code).

Source Files	3449	MVG, McCabe VG Complexity	33954
Directories	1537	C&SLOC, Code and Comment Lines of Code	13661
LOC, Lines of Code	698001	CLOC, Comment Only Lines of Code	191820
BLOC, Blank Lines of Code	93719	CWORD, Commentary Words	1313515
SLOC-P, Physical Executable Lines of Code	412462	HCLOC, Header Comment Lines of Code	74095
SLOC-L, Logical Executable Lines of Code	222623	HCWORD, Header Commentary Words	571732

Figure 1: LocMetrics Count of Source Lines of Code (compiled by the author)

The largest folders with the source data are Breakpad (a crash-reporting library) with 21%, EstEID Tokend (Tokend library for EstEID smartcards) with 16% and X-Road Public (backbone of e-Estonia) and SD-DSS (eSignature creation and validation tool), each covering 14% of the total codes (Figure 2: SLOC-L Distribution).

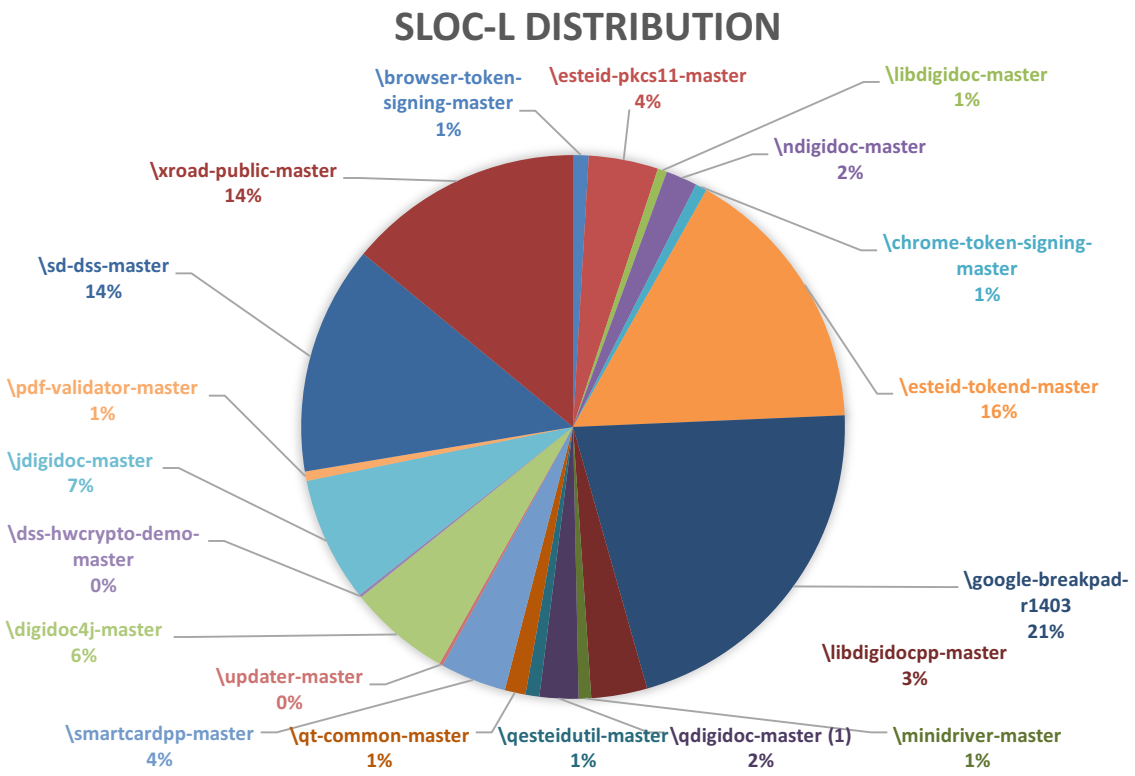


Figure 2: SLOC-L Distribution (compiled by the author)

The programming language should also be considered, as not only can the effort to recode differ, but so can the average salaries (CV-Online OÜ 2016). Over all the repositories, the most prominent languages are C++ and Java (Figure 3).

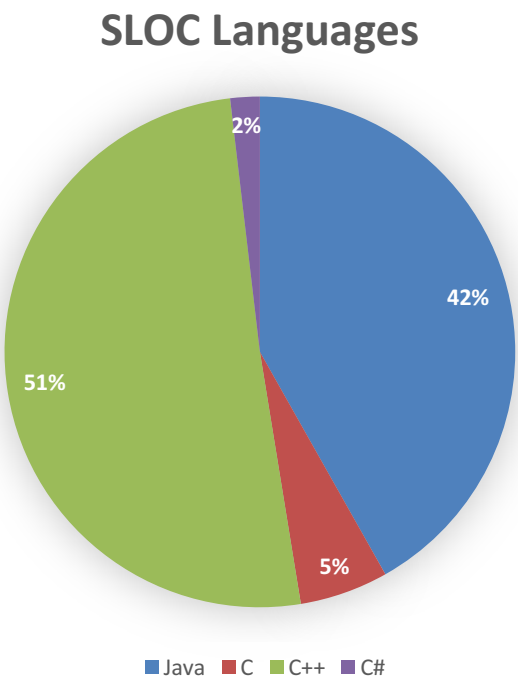


Figure 3: SLOC Languages (compiled by the author)

The following assumptions for the source code are made for larger software developments and will also be applied herein (USC 2000):

1) REVL (% of code thrown away due to requirements evolution and volatility) – **10%**

2) DM (Design Modified) – **10%** (minimal design change to accommodate data placement into blocks)

3) CM (Code Modified) – **25%** (moderate code change to account for re-design of security elements and identity verification)

4) IM (Integration Required) – **50%** (integration to devices largely similar to existing)

5) Assessment and Assimilation (AA) – **6%** - considerable module Test and Evaluation (T&E), documentation

6) Software understanding (SU) – **30 %** - nominal understanding.

7) Programmer unfamiliarity (UNFM) – **0,4** – somewhat familiar.

3.2. Determining Scale Factors and Effort Multipliers

Based on the characteristics provided in Appendix 1, each scale and factor will be assigned with a rating for all calculative considerations in this thesis (Figure 4).

SCALE FACTORS		COST DRIVERS							
		Product		Personnel		Platform		Project	
Scale Factor	**Rating**	**Cost Driver**	**Rating**	**Cost Driver**	**Rating**	**Cost Driver**	**Rating**	**Cost Driver**	**Rating**
PREC	High	RELY	High	ACAP	High	TIME	Nominal	TOOL	Nominal
FLEX	Very High	DATA	High	PCAP	High	STOR	Nominal	SITE	Very High
RESL	Nominal	CPLX	Nominal	PCON	Nominal	PVOL	Nominal	SCED	Low
TEAM	Nominal	RUSE	Very High	APEX	High				
PMAT	Nominal	DOCU	Nominal	PEXP	Nominal				
				LTEX	Nominal				

Figure 4: COCOMO II Scale Factor and Cost Driver ratings (compiled by the author)

4. Results and Discussions

4.1. Source Lines of Code

Based on Formula ASLOC (Adapted Source Lines of Code) = AAM × Reused Volume (USC 2000),

AAM= AA+AAF[1+(0.2×SU×UNFM)],

where

AAF=[(0.4 × DM) + (0.3 × CM) + 0.3 × IM)]

and

Reused Volume (RV) = SLOC + REVL

Therefore,

AAF = [(0.4×10) + (0.3×25) + (0.3×50)] = 26.5 %

Which gives AAM= 0.06+0.265×[1+(0.02×0.3×0.4)] = 0.325636

Because for this project it is estimated that a total of 10% of all code will be reused, the TOTAL Reused Volume = 222623 + 10% = 244885, however since each language is calculated as a separate module, the same should be done for ASLOC volume as well, as COCOMO II rounds up only to full lines of code (USC 2000). Based on the total SLOC counts per language (Figure 3), the following breakdown will be used for calculations:

Language	**SLOC**	**Reused Volume (RV)**
Java	93108	102419
C	12490	13739
C++	112872	124159
C#	4153	4568
Total	222623	244885

Figure 5: Number of reused volume of SLOC's by language

$ASLOC_{Java} = 0.325636 \times 102419 = 33351$

$ASLOC_{C} = 0.325636 \times 13739 = 4473$

$ASLOC_{C++} = 0.325636 \times 124159 = 40430$

$ASLOC_{C\#} = 0.325636 \times 4568 = 1487$

Thus,

$ASLOC_{TOTAL} = 33351 + 4473 + 40430 + 1487 = 79\ 741$

4.2. Exponent Factor E

Because exponent factor E has a formula of (USC 2000):

$$E = B + 0.01 \times \sum_{j=1}^{5} SF_j$$

where $B = 0.91$ (for COCOMO II.2000)

And previously determined ratings for Scale Factors establish the following multipliers Appendix 1.2):

As,

PREC $SF_1 = 2.48$

FLEX $SF_2 = 1.01$

RESL $SF_3 = 4.24$

TEAM $SF_4 = 3.29$

PMAT $SF_5 = 4.68$

then

$E = 0.91 + 0.01 \times (2.48 + 1.01 + 4.24 + 3.29 + 4.68) = 1.067$

Because $E > 1.0$, the project experiences slight diseconomies of scale (USC 2000).

4.3. Effort Multipliers

COCOMO II uses a multiplication of all Effort Multipliers (EM) that influence the effort made, therefore:

EM_{16} = 1.10×1.14×1×1.15×1×0.85×0.88×1×0.88×1×1×1×1×1×1×0.86=0.81635

EM^{17} or SCED applies for the whole project, and therefore will be accounted for separately in the Schedule Estimation chapter (USC 2000), however

EM_{17}=0.81635×1.14 = 0.9306

4.4. Schedule Estimation

The model foresees the schedule calculation as the amount of nominal schedule calendar time to development ($TDEV_{NS}$), which has the formula of (USC 2000, Yang et al. 2008):

$$TDEV = [C \times (PM_{NS})^{(D+0.2\times(E-B))}] \times \frac{SCED\%}{100}$$
$$\text{where } C = 3.67, D = 0.28, B = 0.91$$

The SCED%, at a 'low' rating is measured to be 85% of the nominal schedule. However with the improvements in software developments since 2000, it is estimated for the value at present day to be off by 3% on average in actual projects (Yang et al. 2008). In case of the 'low' rating, the SCED% should be decreased by 3% (Yang et al. 2008), making it a value of 82%.

The formula includes nominal schedule Person Months (PM_{NS}) for effort, which is calculated as (USC 2000):

$$PM_{NS} = A \times Size^{E} \times \prod_{i=1}^{n} EM_i$$
$$\text{where } E = B + 0.01 \times \sum_{j=1}^{5} SF_j$$

,

where 1 Person Month = 152 hours per month (USC 2000).

As A=2.94, B=0.91, C=3.67 and D=0.28 for COCOMO II calibration (USC 2000),

and Size is expressed in KSLOC (Thousands of Source Lines of Code) and is equal to

KSLOC = 79.741

then,

$PM_{NS} = 2.94 \times (79.741)^{1.067} \times 0.9306 = 292.6$

Consequently,

$TDEV_{NS} = 3.67x(292.6)^{(0.28+0.2x(1.067-0.91))} \times 0.82 = 3.67x(292.6)^{0.3114} \times 0.82 = 17.6$

Since the Total effort required (PM_{NS}) is 292,6 and the time to develop ($TDEV_{NS}$) is 17.6 months, that means it would take 292.6/17.6=16.7 people 17.6 months to complete the project.

4.5. Effort Costs

In order to calculate the effort for each module, it is necessary to consider each modules' contribution to the overall effort:

$$PM_{Basic(i)} = PM_{Basic} \times \left(\frac{Size_i}{Size_{Aggregate}} \right)$$

Therefore,

$PM_{Java} = 292.6 \times (33351/79741) = 122.3578$

$PM_{C} = 292.6 \times (4473/79741) = 16.4105$

$PM_{C++} = 292.6 \times (40430/79741) = 148.3292$

$PM_{C\#} = 292.6 \times (1487/79741) = 5.4554$

The average gross salaries in Estonia for programmers in the languages relevant to this case study are 1606 EUR per month for a Java programmer, 1434 EUR/month for C, 1568 EUR/month for C# and 1813 EUR/month for C++ programmer (CV-Online OÜ 2016). Considering the total out of pocket costs for the employer (i.e. the salary fund, roughly an additional 33.8% from social and unemployment taxes), the average salary of each programmer and the total effort in Person Months (PM), the total costs of each module are:

Java – 2148 EUR × 122.3578 = 262 825 EUR

C – 1919 EUR × 16.4105 = 31 492 EUR

C++ – 2425 EUR × 148.3292 = 359 698 EUR

C# - 2097 EUR × 5.4554 = 11 440 EUR

Using the COCOMO II software development cost estimation tool, the total baseline cost of recoding X-Road and Estonian Electronic Identity Software to blockchain, would then in sum be: Solution Cost (SC)= 262 825 + 31 492 + 359 698 + 11 440 = 665 455 EUR.

4.6. Scenarios

Some cost drivers and scale factors can affect the total costs more significantly than others (Appendix 2). As seen from Appendix 2, underestimating the pressure of meeting the scheduled deadline for example can increase the costs of the project by over 25%. These factors however help to recognize most the influential cost drivers and identify the potential for cost-saving.

4.6.1. Scenario A:

The most significant cost savings would derive from hiring higher-capability analysts (Appendix 2). Bringing the capability of the team of analysts (ACAP) from the current 75th percentile (high) to 90th percentile (very high) would cut the costs of the whole project by over 16%, or just under 110 000 EUR. Naturally, extra training for staff would introduce some new occurring costs as well. There are a variety of online courses available from prominent universities such as MIT, Princeton University, Duke University and Stanford, that introduce the basics of blockchain (Holmes 2015). For a 10 week course, the price tag ranges from $3200-3600 (Holmes 2015). There are also free courses available from Bitcoin University and Coursera (Coursera 2016, Blockchain University 2016).

Sending a 3-member team of analysts to 2 courses prior to starting the project would end up costing around 22 000 USD or 19 000 EUR in course fees + 10 000 EUR in flights/accommodation. Yet, assuming the course can bring the capability of the personnel to the required 90th percentile, it will still lower the total cost of the software development by 81 000 EUR and bring the total cost down to 584 455 EUR.

4.6.2. Scenario B

The second most effective way to reduce costs is to increase the capability of the programmers involved in the project (PCAP) (Appendix 2). Since Java and C++ are overwhelmingly the dominant languages used in the project, 10 programmers (Six C++ and four Java programmers) are assumed to receive access to the same trainings as in Scenario A. Therefore, the cost of training fees for the necessary 10 programmers would be ($3500 × 2 courses × 10 people) = $70 000 or about 60 000 €. With an additional 20 000 EUR for flights and accommodation, total out of pocket costs for such training would be 80 000 EUR

The respective savings from increasing the capability of the programmers from high (75th percentile) to very high (90th percentile) would be:

	Baseline	PCAP very high	Difference
Java	262 825 €	226 976 €	-35 849 €
C++	359 698 €	310 635 €	-49 063 €
Total			**-84 912 €**

Figure 6: Cost savings from PCAP increase on Java and C++ modules to 90th percentile (composed by the author).

Therefore implementing training only for programmers would cost about 80 000 €, but reduce the overall costs by only 4 912 €.

4.6.3. Scenario C

Considering savings from both Scenario A and Scenario B, it would make sense to assume that it would be beneficial to have both the necessary programmers (10) and the team of analysts (3) receive extra training in blockchain programming and software development to bring the capabilities of both teams to 90th percentile.

The training costs for both analysts and programmers will a total of (based on calculations in Scenario A and Scenario B) 109 000 EUR, however the total project costs would go down significantly:

	Baseline	PCAP and ACAP very high	Difference
Java	262 825 €	189 608 €	-73 217 €
C	31 492 €	26 306 €	-5 186 €
C++	359 698 €	259 494 €	-100 204 €
C#	11 440 €	9556 €	-1 884 €
Total	**665 455 €**	**484 964**	**-180 491 €**

Figure 7: Cost savings from increase of PCAP on Java and C++ modules and ACAP to 90th percentile (composed by the author).

Still, considering the initial spending for training (109 000 €), the total savings from Scenario C would be 180 491 € - 109 000 € = 71 491 €, with total project costs at 593 964 €.

Comparing all possible scenarios (Figure 9), Scenario A, where only investments to increase ACAP were made, reduces the overall costs of the project the most.

	Costs	**Reduction in costs**
Baseline	665 455 €	-
Scenario A	584 455 €	81 000 €
Scenario B	660 543 €	4 912 €
Scenario C	593 964 €	71 491 €

Figure 8: Comparison of cost savings from increased ACAP and PCAP ratings (composed by the author).

5. Return On Security Investment

Solution Costs and calculations are provided in chapter 4. Results and Discussions.

In order to determine the Annual Loss Expectancy (ALE), the values of Annualized Rate of Occurrence (ARO), Exposure Factor (EF) and Asset Value (AV) should also be concluded, as:

ALE = SLE × ARO

and

SLE = AV × EF.

Firstly, to find the Single Loss Expectancy (SLE), a total value of the asset should be determined. Because of the scale of the program and its budgetary distribution amongst several different departments of the government, such as the Information System Authority (RIA), IT and Development Centre of the Ministry of the Interior and Certification Centre Ltd., it is rather complex to give an accurate measurement for the value of the given asset – the X-road technology and Electronic Identity Software, more so because some benefits, such as savings on time, paper and fuel are deemed priceless.

***Figure 10**: RIA cost and benefit analysis* (Kiivet 2015)

Costs	
Creation of the document	€ 36M
Electronic certification	€11M
Issuing	€9M
Base software	€5M
User-support	€3M
Support in applications	€25M
TOTAL	**€89 M**
Benefits	
State fees	€40M
Cost savings on digital vs physical signature	€251M
Cost savings on identification	€196M
Paper and fuel savings	priceless
TOTAL	**€487 M**

In 2013, RIA ran a cost and benefits analysis on the whole digital identification and e-signatures service for the upcoming 10 years (Figure 10: RIA cost and benefit analysis). According to their own estimations, over the given time, the services would cost €89 M and bring in benefits in the amount of €487 M, excluding any benefits from saving natural

resources. The value may be even higher as of the present day due to unexpected success of the e-services. In 2015, the head of the IT-department for Estonian Health Insurance Fund (Haigekassa) Raimo Laus stated that the use of e-prescriptions is facilitating a reduction in costs of about 22M € per annum (Matson 2015).

If a significant attack occurs on the identification process, the core trust of the digital signature process will suffer, potentially even inducing a switch back to physical services. Based on the report, the value of the electronic identification and digital signature system can be expressed through the benefit of having it in place instead of physical identification and signing process. Hence,

ASSET VALUE = €487M – €89M = €398M over the span of 10 years.

Exposure factor is calculated by subtraction from full 100% of various characteristics and their value (e.g. 30% will be deducted for having redundancies/back-up copies of the files, 5% will be deducted due to the presence of a firewall etc.) (Tan 2002). However, because one of the major inputs in calculating ROSI, that the SLE is dependent on the Exposure Factor and exposure depends on the type of attack, time to discovery, security measures at the target department and various other aspects, a matrix with a full scale of exposure will be provided in this thesis, to better understand the impact of the investment in case of all types of attacks.

X-Road was first introduced in 2001 and hasn't experienced any major security breaches in the 15 years of its operation. The series of cyber attacks in 2007 were mainly DDOS (Distributed Denial of Services) targeting websites, rather than the X-road itself (Nazario 2007). Several precautionary measures were introduced after the attacks, to further improve the security of the system. Regardless, it would only take one similar attack on X-road to significantly compromise the services. Because the SLE assumes the value of the project for 10 years, it would be pertinent to also assume a potential attack within those ten years for the purposes of calculating the ROSI.

Thus, the Annualized Rate of Occurrence (ARO) will be 1/10 or 10%.

The last variable in the formula, the % of risk mitigated, is similarly to the exposure factor very much dependent on the scale and type of the attack and the target database. Furthermore, the last variable is implicated with the extent of how the blockchain method

is applied and dependent on the success of being able to predict the weakest security factors in the data processing chain. In case of a smaller cyberattack such as DoS (Denial of Service) to a small database, the costs (or rather lack of benefits during the off-time) may not be significant, however in case delicate data in vast quantities is stolen, the impact is significantly higher. Therefore, various values of the % of Risk Mitigated will also be reflected further on in the thesis.

For example, the ROSI for Scenario A provided in section 4.6 Approaches, with an Exposure Factor of 30% and Risk Mitigated of 85% (as described by Sonnenreich (2006: 61)), is based on the ROSI formula provided in 2. Literature Review as follows:

$$ROSI = \frac{(ALE \times 0.85) - 584\,455}{584\,455}$$

As

ALE = SLE × ARO

and

SLE = AV × EF,

then,

ALE = 398 000 000 × 0.3 × 0.1 = 11 940 000

and thus,

$$ROSI = \frac{(11\,940\,000 \times 0.85) - 584\,455}{584\,455} = 16.364 = 1636.5\ \%$$

The results in Table 1 indicate that, when considering the whole X-road platform and the electronic identity software, the investment in security software that could prevent even a smaller attack would bring in meaningful returns on investment. Only in cases when both Exposure Factor (EF) and % of Risk Mitigated are 10% or lower would the investment yield negative return on investment.

Assuming adequate research into the rebuilding process is done and the security solution works as intended, a fixed value of 85% for Risk Mitigated can be assigned, as proposed by Sonnenreich (2006: 61). As mentioned prior, the value of the asset is difficult to estimate and the initial data involves all of the usage value across Electronic Identity Software and X-road implementation. Therefore, in order to provide an authentic perception on potential ROSI, the asset value impact on the ROSI will be observed (Table 2)

Table 1. Return on Security Investment calculation matrix (composed by the author)

ROSI (%)		Exposure Factor									
		10.00%	**20.00%**	**30.00%**	**40.00%**	**50.00%**	**60.00%**	**70.00%**	**80.00%**	**90.00%**	**100.00%**
% of Risk Mitigated	**10%**	-31.9%	36.2%	104.3%	172.4%	240.5%	308.6%	376.7%	444.8%	512.9%	581.0%
	20%	36.2%	172.4%	308.6%	444.8%	581.0%	717.2%	853.4%	989.6%	1125.8%	1262.0%
	30%	104.3%	308.6%	512.9%	717.2%	921.5%	1125.8%	1330.1%	1534.3%	1738.6%	1942.9%
	40%	172.4%	444.8%	717.2%	989.6%	1262.0%	1534.3%	1806.7%	2079.1%	2351.5%	2623.9%
	50%	240.5%	581.0%	921.5%	1262.0%	1602.4%	1942.9%	2283.4%	2623.9%	2964.4%	3304.9%
	60%	308.6%	717.2%	1125.8%	1534.3%	1942.9%	2351.5%	2760.1%	3168.7%	3577.3%	3985.9%
	70%	376.7%	853.4%	1330.1%	1806.7%	2283.4%	2760.1%	3236.8%	3713.5%	4190.2%	4666.8%
	80%	444.8%	989.6%	1534.3%	2079.1%	2623.9%	3168.7%	3713.5%	4258.2%	4803.0%	5347.8%
	90%	512.9%	1125.8%	1738.6%	2351.5%	2964.4%	3577.3%	4190.2%	4803.0%	5415.9%	6028.8%
	100%	581.0%	1262.0%	1942.9%	2623.9%	3304.9%	3985.9%	4666.8%	5347.8%	6028.8%	6709.8%

Table 2*:* ROSI with assumed 85% risk mitigation (composed by the author).

			Exposure Factor									
		Value (M€)	**10.00%**	**20.00%**	**30.00%**	**40.00%**	**50.00%**	**60.00%**	**70.00%**	**80.00%**	**90.00%**	**100.00%**
Value ASset	**50%**	**199**	189.4%	478.8%	768.2%	1057.7%	1347.1%	1636.5%	1925.9%	2215.3%	2504.7%	2794.1%
	60%	**238.8**	247.3%	594.6%	941.9%	1289.2%	1636.5%	1983.8%	2331.1%	2678.4%	3025.7%	3373.0%
	70%	**278.6**	305.2%	710.4%	1115.5%	1520.7%	1925.9%	2331.1%	2736.3%	3141.4%	3546.6%	3951.8%
	80%	**318.4**	363.1%	826.1%	1289.2%	1752.3%	2215.3%	2678.4%	3141.4%	3604.5%	4067.6%	4530.6%
	90%	**358.2**	420.9%	941.9%	1462.8%	1983.8%	2504.7%	3025.7%	3546.6%	4067.6%	4588.5%	5109.5%
	100%	**398**	478.8%	1057.7%	1636.5%	2215.3%	2794.1%	3373.0%	3951.8%	4530.6%	5109.5%	5688.3%
	110%	**437.8**	536.7%	1173.4%	1810.1%	2446.9%	3083.6%	3720.3%	4357.0%	4993.7%	5630.4%	6267.1%
	120%	**477.6**	594.6%	1289.2%	1983.8%	2678.4%	3373.0%	4067.6%	4762.2%	5456.8%	6151.4%	6846.0%
	130%	**517.4**	652.5%	1405.0%	2157.4%	2909.9%	3662.4%	4414.9%	5167.4%	5919.8%	6672.3%	7424.8%
	140%	**557.2**	710.4%	1520.7%	2331.1%	3141.4%	3951.8%	4762.2%	5572.5%	6382.9%	7193.3%	8003.6%
	150%	**597**	768.2%	1636.5%	2504.7%	3373.0%	4241.2%	5109.5%	5977.7%	6846.0%	7714.2%	8582.4%

The calculations from Table 2 clearly indicate that as long as a reasonable assumption can be placed on the functionality of the technology, even in case of a single attack with a low exposure rate, the transition provides high rates of ROSI.

Because ROSI is often evaluated on a smaller scale, it is difficult to draw appropriate comparisons from scientific literature. Most of the time, the ROSI % remains within the range of few hundreds of percentages (Sonnenreich: 2006, 46, ENISA 2012: 5), however because of the scale and sensitivity (civil liability) of the given project, it is logical to assume larger result on ROSI as well. The investment to be made is connected to the size of the platforms and software handling data and not necessarily to the data itself, however losses are accounted for based on the loss of data or denial of services. Hence, there's a certain scalability already built into the investment.

These result seem rather high for a regular ROI, however this is mostly the case in security investments, when it only takes one coordinated attack to make the value of the asset plummet (Kolochenko 2015). It's also noteworthy to mention the notion of Gordon & Loeb (2002: 122) that the optimal amount to spend on security investment should never exceed 37% of the potential loss and that in fact it is rarely ever the case. Based on this statement and the highly favorable ROSI , it is deeply recommended to transfer X-road and eID systems onto blockchain.

6. Conclusion

The purpose of this thesis was to provide a first real case scenario of transfer to blockchain technology for database management, to estimate the development cost and the return on security investment (ROSI).

This thesis analyzed the potential costs of transition of a database to blockchain, which in itself is a very poorly researched subject. Understandably, with the development of blockchain technology and improvement in cost calculation models, the accuracy will improve and this thesis sheds light on a dimension of the research which has not been shown before.

Transfer to blockchain is still a work in progress and hence not much scientific literature is available, particularly with real-life examples. There is a strategy proposed by Ethereum (Buterin 2013) to develop blockchain applications which helps to transition from the current code to blockchain. Many experts agree that Ethereum is a solid base on which to build the transfer, but also admit the unprecedented scalable examples (Czepluch et al. 2015).

Software cost estimation has been a well practiced occupation, however it still lacks accuracy despite many different estimation models such as SLIM, Checkpoint, PRICE-S, SEER and COCOMO. The latter has been by far the most dominant, especially with the next generation model COCOMO II (Jørgensen & Shepperd 2007). The model uses the source code and a variety of scale factors and cost multipliers that help to shape the estimation of total effort that would go into the particular development. It then accounts for the schedule implications and local labor fees to estimate the total cost of development (USC 2000).

Despite numerous models for calculating ROSI, many of them also struggle with scalability or have to rely on sets of historic data (Rosenquist 2007, Sonnenreich 2006). It is also noted that investing in error detection and eliminating mistakes in code writing in the early stages will help to save the costs of development (Kramer et al. 2015).

The thesis assesses various scenarios for transfering the Estonian Electronic Identity Software and the data communication platform X-road to blockchain and estimates that

with certain investments into the capability of the programmers (participation in training) before initiating the project, the total cost of the effort would be 437 515 EUR.

It then provides calculative proof that, in the majority of the cases, the investment would be worthwhile, as only few incidences where the potential attack would have a very low exposure rate (around 10%) will yield a negative return on the investment. Furthermore, similar conclusions can be drawn, even if the value of the asset is 50% of the total value assumed by RIA itself, providing a great example of scalability for calculating ROSI.

This thesis provides the first case of considering technology transfer costs in the case of transitioning to blockchain technology and studies financial gains (ROSI) for the consideration. It also estimates ROSI on a very large governmental project, thus improving the level of scalability for both software development costs and ROSI models. Additionally, it provides a sense of cost implication of security investments for RIA, the Estonian Information System Authority, and hopefully some encouragement to consider novel technologies and justifications for large scale investments for security considerations.

The study also has some limitations. The actual cost of maintaining the digital identification platform and e-government programs are hard to trace, because of their span over several ministries and organizations. Thus it is also difficult to estimate the true value of the asset. Similarly, the study operates under the assumption that blockchain technology is indeed as big an improvement in the security and transparency as many estimate it to be. Despite several papers and studies that theorize the usage and benefits of blockchain technology and success in cryptocurrencies or even small scale data management applications, the true tests of the technology are still ahead. Organizations such as Ethereum, SAFEnetwork and Open Garden have put forward some very promising applications for the use of blockchain technology, but until the usage rates go higher and they are able to operate on very large scales, it is impossible to tell for sure whether the technology is applicable for such pan-national usage.

References

1. **Al-Humaigani, M., and D.B. Dunn**. 2003. "A MODEL OF RETURN ON INVESTMENT FOR INFORMATION SYSTEMS SECURITY." *IEEE International Symposium.* 483-485.

2. **Back, A., Corallo, M., Dashjr, L., Friedenbach, M., Maxwell, G., Miller, A., Timón, J.P, and Wuille P**. 2014. *Enabling Blockchain Innovations with Pegged Sidechains.* Blockstream.

3. **Baraniuk, C.** 2016. "Government urged to use Bitcoin-style digital ledgers." *BBC News*, 19 January.

4. **Berson, and Zemlok**. 1995. Apparatus for verifying an identification card and identifying a person by means of a biometric characteristic. USA Patent US 5469506 A. 21 November.

5. **Bijker, W.E, and Pinch, T.**. 1987. *The Social Construction of Technological Systems.* Massachussets Institute of Technology.

6. **Bissias, G., Ozisik, A.P., ‘,Levine B., and Liberatore M.** 2014. *Sybil-Resistant Mixing for Bitcoin.* School of Computer Science, University of Massachusetts Amherst.

7. **Blockchain University**. n.d. *Blockchain University Courses.* Accessed March 17, 2016. http://blockchainu.co/upcoming/.

8. **Bloomberg, J**. 2016. "Something Rotten In The State Of Bitcoin." *Forbes*, 18 January.

9. **BNS**. 2015. "Estonia to launch e-governance project with Georgia." *The Baltic Times*, 9 November.

10. **Boehm, B., Abts C., and Chulani, S.**. 2000. *Software development cost estimation approaches - A survey.* LA: USC.

11. **Boehm, B., Clark, B., Horowitz, E., Westland, C., Madachy, R., and Selby, R.**. 1995. "Cost Models for Future Software Life Cycle Processess: COCOMO 2.0." *Annals of Software Engineering* (Baltzer Science Publishers) 1: 57-94.

12. **Böhme et al**. 2015. "Bitcoin: Economics, Technology, and Governance." *Journal of Economic Perspectives* 29 (2): 213-238.

13. **Booz, Allen, Hamilton**. 2014. *Cyber ROI: A practical approach to quantifying the financial benefits of cybersecurity.* Ponemon institute .

14. **Brodbeck, L.**. 2015. "Is Bitcoin The Next Internet." *Yahoo Finance*, 26 February.

15. **Buterin, V**. 2013. *Ethereum White Paper: A NEXT GENERATION SMART CONTRACT & DECENTRALIZED APPLICATION PLATFORM.* ethereum.org, 14-21.

16. **Buterin, V**. 2015. *A Next Generation Smart Contract & Decentralized Application Platform.* Ethereum.

17. **Buterin, V**. 2016. *Privacy on the Blockchain.* 15 January. Accessed April 7, 2016. https://blog.ethereum.org/2016/01/15/privacy-on-the-blockchain/.

18. **Carter, L., and Bélanger, F.**. 2005. "The utilization of e-government services: citizen trust, innovation and acceptance factors." *Info Systems J* (Blackwell Publishing Ltd) (15): 5-25.

19. **Cavusoglu, H., Mishra B., and Raghunathan S.,**. 2004. "A Model for Evaluating IT Security Investments." *COMMUNICATIONS OF THE ACM* 47 (7).

20. **Coursera**. n.d. *Bitcoin and Cryptocurrency Technologies.* Accessed March 17, 2016. https://www.coursera.org/course/bitcointech.

21. **CV-Online OÜ**. 2016. *Palgad kategoorias: Infotehnoloogia (IT).* Palgad.ee.

22. **Czepluch, J.S., Lollike, N.Z. and Malone, S.O.** 2015. "The Use of Block Chain Technology in Different Application Domains." The IT University of Copenhagen, Copenhagen.

23. **e-Estonia.com**. 2015. 25 October. Accessed March 18, 2016. https://e-estonia.com/japan-is-implementing-the-id-card-following-estonia-example/.

24. **Enikeieff et al**. 1965. Electronic identification system employing a data bearing identification card. USA Patent US 3221304 A. 30 November.

25. **ENISA**. 2012. *Introduction to Return on Security Investment.* The European Network and Information Security Agency (ENISA).

26. **ERR**. 2013. *Finland Follows Estonia's Footsteps in e-Governance.* 2 February. Accessed March 18, 2016. http://news.err.ee/v/news/scitech/664fbc95-f5d1-4c3b-a2be-fd735c673b9e/finland-follows-estonias-footsteps-in-e-governance.

27. **Holmes, B**. 2015. *Bitcoin and Blockchain Education: The Key to Innovation.* Brave NewCoin, 08 September .

28. **Jørgensen, M., and Shepperd, M.**. 2007. "Systematic Review of Software Development Cost Estimation Studies." *IEEE TRANSACTIONS ON SOFTWARE ENGINEERING* 33 (1).

29. **Kalja, A., Reitsakas, A. and Saard, N.**. 2005. *eGovernment in Estonia: Best Practices.* Tallinn: Inst. of Cybernetics at Tallinn Univ. of Technology.

30. **Kiivet, H.**. 2015. *Koosvõime lahenduste talituse juhataja/CO* (17 November).

31. **Kitsing, M**. 2011. "Success Without Strategy: E-Government Development in Estonia." *Policy & Internet*, Art 5.

32. **Kolochenko, I.**. 2015. *How to calculate ROI and justify your cybersecurity budget.* CSO.

33. **Kosba, A., Miller, A., Shi, E., Wen, Z., Papamanthou, X**. 2015. *Hawk: The Blockchain Model of Cryptography and Privacy-Preserving Smart Contracts.* University of Maryland and Cornell University, Cryptology ePrint Archive, Report 2015/675.

34. **Kramer et al.** 2015. "INCREASE on Investment of Software Development Life Cycle RETURN by Managing the Risk —A Case Study." *Defense ARJ* (Defense Acquisition University,9820 Belvoir Rd Ste 3,Fort) 22 (2): 174-191.

35. **Krause, M. and Tipton, H.F**. 1993. *Handbook of Information Security Management.* CRC Press LLC.

36. **Lane, W.F.** 1997. Self-authenticating identification card with fingerprint identification . Patent US 5623552 A. 22 April.

37. **Lavinskaya, A.** 2016. "Where to use blockchain besides finance?" *Coinfox.*

38. **Lawrence, A.G. and Loeb, M.P**. 2002. "The Economics of Information Security Investment." University of Maryland.

39. **Lederer, A.L., and Prasad, J.**. 1995. *Causes of Inaccurate Software Development Cost Estimates.* New York: Elsevier Science Inc.

40. **Mansel, T.**. 2013. "How Estonia became E-stonia." *BBC News*, 16 May.

41. **Matson, A.,** 2015. “Digiretsept säästab inimestele 22 miljonit eurot.” *Äripäev*, February 19th

42. **Matthews, C.**. 2014. "Bit Con? Veteran fraud expert sets his sights on bitcoin." *Fortune*, 24 October.

43. **Mougayar, W.**. 2014. *The Blockchain is the New Database, Get Ready to Rewrite Everything.* December 27. http://startupmanagement.org/2014/12/27/theblockchain-is-the-new-database-get-ready-to-rewriteeverything/.

44. **Nakamoto, S.**. 2008. *Bitcoin: A Peer-to-Peer Electronic Cash System.* www.bitcoin.org.

45. **Naughton, J.** 2016. "Is Blockchain the most important IT invention of our age?" *The Guardian*, 24 January.

46. **Nazario, J**. 2007. *Estonian DDoS Attacks – A summary to date.* Arbor Networs - DDos Experts.

47. **Nelson, E.A**. 1966. "MANAGEMENT HANDBOOK FOR THE ESTIMATION OF COMPUTER PROGRAMMING COSTS." Electronic Systems Division, Air Force Systems Command, United States Air Force.

48. **Neumann, L**. 2013. "Security Challenges of Current Federated eID Architectures." In *ISSE 2013 Securing Electronic Business Processes*, by Reimer et al., 21-32.

49. **Nguyen, V., Deeds-Rubin, S., Tan, T., and Boehm, B.** 2007. *A SLOC Counting Standard.* University of Southern California.

50. **Nixon, P. G**. 2007. *E-government in Europe: Re-booting the State.* Routledge.

51. **Plassaras, N.A**. 2013. "Regulating Digital Currencies: Bringing Bitcoin within the Reach of the IMF." *Chicago Journal of International Law* 14 (1): 377-407.

52. **Ponemon Institute and IBM**. 2014. *2014 Cost of Data Breach Study.* Ponemon Institute.

53. **Popper, N**. 2015. "Bitcoin Basics." *New York Times*, 4 November.

54. **Population Register Center**. 2016. *GitHub Repository.* Helsinki: Population Register Center.

55. **PWC**. 2015. *2015 INFORMATION SECURITY BREACHES SURVEY.* HM Government.

56. **Riigi Infosüsteemi Amet (RIA)**. 2016. *Open Electronic Identity.* Tallinn: RIA, 16 March.

57. **Rosenquist, M. 2007**. *Measuring the Return on IT Security Investments.* White Paper, Intel Corporation, 5.

58. **Schware, R. and Deane, A.**. 2003. *Deploying e-government programs: the strategic importance of "I" before "E".* Global Information and Communication Technologies Department, The World Bank.

59. **Scott, M.**, 2014. "Estonians Embrace Life in a Digital World." *New York Times*, 08 October.

60. **Sonnenreich, W.**. 2006. *Return On Security Investment (ROSI) - A Practical Quantitative Model.* Australian Computer Society Inc.

61. **Stein, T.** 1989. Electronic funds transfer system with means for verifying a personal identification number without pre-established secret keys . USA Patent US4797920 A. 10 January.

62. **Steinber, J., and Tipton, H.F,**. 2011. *Official (ISC)2 Guide to the ISSMP CBK*. CRC Press.

63. **Stuntz, J.**, 2014. *S 2 ERC Project: A Review of Return on Investment for Cybersecurity.* Georgetown University Security and Software Engineering Research Centre.

64. **Swan, M**., 2015. *Blockchain: Blueprint for a New Economy.* O'Reilly Media.

65. **Swanson, T**. 2015. *Consensus-as-a-service: a brief report on the emergence of permissioned, distributed ledger systems.* R3 CEV .

66. **Tan, D.** 2002. *Quantitative Risk Analysis Step-By-Step.* SANS Institute InfoSec Reading Room.

67. **Tapscott, D.** 2015. "What's the Next-Generation Internet? Surprise: It's All About the Blockchain!" *Huffington Post*, 12 March.

68. **The Economist**. 2015. "The great chain of being sure about things." *The Economist.*

69. **The Economist**. 2015. "The next big thing." *The Economist.*

70. **The Economist**. 2015. "The promise of the blockchain: The trust machine." *The Economist.*

71. **Trent et al**. 2016. "BigchainDB: A Scalable Blockchain Database." ascribe GmbH, Berlin.

72. **Tsalis, N., Theoharidou, M., and Gritzalis, D**. 2013. "Return on Security Investment for Cloud Platforms." *2013 IEEE International Conference on Cloud Computing Technology and Science.* IEEE Computer Society. 132-137.

73. **Tsiakis, T.K, and Pekos, G.D.**. 2008. "Analysing and determining Return on Investment for Information Security ." *International Conference on Applied Economics* . 879-883.

74. **UK Office for Science**. 2015. "Distributed Ledger Technology: beyond block chain."

75. **USC.** 2000. *COCOMO II - Model Definition Manual.* Center for Software Engineering, University of Southern California.

76. **Vigna, P., and Casey M.,**. 2015. *The Age of Cryptocurrency: How Bitcoin and Digital Money Are Challenging the Global Economic Order.* St. Martin's Press.

77. **West, D.M. 2004**. "E-Government and the Transformation of Service Delivery and Citizen Attitudes." *Public Administration Review* 64 (1): 15-27.

78. **Yang, Y, Chen Z., Valerdi, R., and Boehm**, **B**. 2008. *Effect of Schedule Compression on Project Effort.* Los Angeles.

Appendix 1

Appendix 1.1 COCOMO II Scale factors (COCOMO II User manual: pp 74)

Cost Drivers	Very Low	Low	Nominal	High	Very High	Extra High
RELY	slight inconvenience	low, easily recoverable losses	moderate, easily recoverable losses	high financial loss	risk to human life	
DATA		Testing DB bytes / Pgm SLOC < 10	10 ≤ D/P < 100	100 ≤ D/P < 1000	D/P > 1000	
CPLX	see Table 19					
RUSE		none	across project	across program	across product line	across multiple product lines
DOCU	Many life-cycle needs uncovered	Some life-cycle needs uncovered.	Right-sized to life-cycle needs	Excessive for life-cycle needs	Very excessive for life-cycle needs	
TIME			≤ 50% use of available execution time	70%	85%	95%
STOR			≤ 50% use of available storage	70%	85%	95%
PVOL		major change every 12 mo.; minor change every 1 mo.	major: 6 mo.; minor: 2 wk.	major: 2 mo.; minor: 1 wk.	major: 2 wk.; minor: 2 days	
ACAP	15th percentile	35th percentile	55th percentile	75th percentile	90th percentile	
PCAP	15th percentile	35th percentile	55th percentile	75th percentile	90th percentile	
PCON	48% / year	24% / year	12% / year	6% / year	3% / year	
APEX	≤ 2 months	6 months	1 year	3 years	6 years	
PLEX	≤ 2 months	6 months	1 year	3 years	6 year	
LTEX	≤ 2 months	6 months	1 year	3 years	6 year	
TOOL	edit, code, debug	simple, frontend, backend CASE, little integration	basic lifecycle tools, moderately integrated	strong, mature lifecycle tools, moderately integrated	strong, mature, proactive lifecycle tools, well integrated with processes, methods, reuse	

Cost Drivers	Very Low	Low	Nominal	High	Very High	Extra High
SITE: Collo-cation	International	Multi-city **and** multi-company	Multi-city **or** multi-company	Same city or metro area	Same building or complex	Fully collocated
SITE: Com-muni-cation	Some phone, mail	Individual phone, FAX	Narrow-band email	Wide-band electronic communica-tion.	Wide-band elect. comm, occasional video conf.	Interactive multimedia
SCED	75% of nominal	85% of nominal	100% of nominal	130% of nominal	160% of nominal	

Appendix 1.2 COCOMO II Scale factors (COCOMO II User manual: pp 73)

Scale Factors	Very Low	Low	Nominal	High	Very High	Extra High
PREC	thoroughly unprecedented	largely unprecedented	somewhat unprecedented	generally familiar	largely familiar	thoroughly familiar
SF_j:	6.20	4.96	3.72	2.48	1.24	0.00
FLEX	rigorous	occasional relaxation	some relaxation	general conformity	some conformity	general goals
SF_j:	5.07	4.05	3.04	2.03	1.01	0.00
RESL	little (20%)	some (40%)	often (60%)	generally (75%)	mostly (90%)	full (100%)
SF_j:	7.07	5.65	4.24	2.83	1.41	0.00
TEAM	very difficult interactions	some difficult interactions	basically cooperative interactions	largely cooperative	highly cooperative	seamless interactions
SF_j:	5.48	4.38	3.29	2.19	1.10	0.00
PMAT	The estimated Equivalent Process Maturity Level (EPML) or					
	SW-CMM Level 1 Lower	SW-CMM Level 1 Upper	SW-CMM Level 2	SW-CMM Level 3	SW-CMM Level 4	SW-CMM Level 5
SF_j:	7.80	6.24	4.68	3.12	1.56	0.00

Appendix 1.3. COCOMO II Scale factors (COCOMO II User manual: pp 75)

Baseline Effort Constants: **A** = 2.94; **B** = 0.91 Baseline Schedule Constants: **C** = 3.67; **D** = 0.28							
Driver	**Symbol**	**VL**	**L**	**N**	**H**	**VH**	**XH**
PREC	SF_1	6.20	4.96	3.72	2.48	1.24	0.00
FLEX	SF_2	5.07	4.05	3.04	2.03	1.01	0.00
RESL	SF_3	7.07	5.65	4.24	2.83	1.41	0.00
TEAM	SF_4	5.48	4.38	3.29	2.19	1.10	0.00
PMAT	SF_5	7.80	6.24	4.68	3.12	1.56	0.00
RELY	EM_1	0.82	0.92	1.00	1.10	1.26	
DATA	EM_2		0.90	1.00	1.14	1.28	
CPLX	EM_3	0.73	0.87	1.00	1.17	1.34	1.74
RUSE	EM_4		0.95	1.00	1.07	1.15	1.24
DOCU	EM_5	0.81	0.91	1.00	1.11	1.23	
TIME	EM_6			1.00	1.11	1.29	1.63
STOR	EM_7			1.00	1.05	1.17	1.46
PVOL	EM_8		0.87	1.00	1.15	1.30	
ACAP	EM_9	1.42	1.19	1.00	0.85	0.71	
PCAP	EM_{10}	1.34	1.15	1.00	0.88	0.76	
PCON	EM_{11}	1.29	1.12	1.00	0.90	0.81	
APEX	EM_{12}	1.22	1.10	1.00	0.88	0.81	
PLEX	EM_{13}	1.19	1.09	1.00	0.91	0.85	
LTEX	EM_{14}	1.20	1.09	1.00	0.91	0.84	
TOOL	EM_{15}	1.17	1.09	1.00	0.90	0.78	
SITE	EM_{16}	1.22	1.09	1.00	0.93	0.86	0.80
SCED	EM_{17}	1.43	1.14	1.00	1.00	1.00	

Appendix 2

Summary of changes by one category per each SF/CD (composed by the author)

Scale Factor/ Cost Driver	Change in Range	Cost	Change %
SCHED	Down	834 746,75 €	125,44%
ACAP	Down	782 907,81 €	117,65%
CPLX	Up	778 582,35 €	117,00%
PVOL	Up	765 273,25 €	115,00%
RELY	Up	762 278,70 €	114,55%
PCAP	Down	756 223,06 €	113,64%
APEX	Down	756 223,06 €	113,64%
DATA	Up	747 172,87 €	112,28%
PCON	Down	745 309,60 €	112,00%
DOCU	Up	738 655,05 €	111,00%
TIME	Up	738 655,05 €	111,00%
LTEX	Down	725 345,95 €	109,00%
PLEX	Down	725 345,95 €	109,00%
TOOL	Down	725 345,95 €	109,00%
SITE	Down	719 623,04 €	108,14%
RUSE	Up	717 560,13 €	107,83%
STOR	Up	698 727,75 €	105,00%
PMAT	Down	679 096,83 €	102,05%
RESL	Down	677 765,92 €	101,85%
PREC	Down	676 301,92 €	101,63%
TEAM	Down	674 971,01 €	101,43%

FLEX	Down	674 372,10 €	101,34%
Baseline		665 455,00 €	100,00%
FLEX	Up	656 737,54 €	98,69%
TEAM	Up	656 005,54 €	98,58%
PREC	Up	654 807,72 €	98,40%
RESL	Up	653 343,72 €	98,18%
PMAT	Up	652 079,35 €	97,99%
RUSE	Down	619 139,33 €	93,04%
SITE	Up	619 006,24 €	93,02%
APEX	Up	612 551,33 €	92,05%
DOCU	Down	605 564,05 €	91,00%
LTEX	Up	605 564,05 €	91,00%
PLEX	Up	605 564,05 €	91,00%
RELY	Down	604 965,14 €	90,91%
PCON	Up	598 909,50 €	90,00%
TOOL	Up	598 909,50 €	90,00%
SCHED	Up	583 737,13 €	87,72%
DATA	Down	583 737,13 €	87,72%
CPLX	Down	578 945,85 €	87,00%
PVOL	Down	578 945,85 €	87,00%
PCAP	Up	574 686,94 €	86,36%
ACAP	Up	555 854,56 €	83,53%
TIME	Down	N/A	N/A
STOR	Down	N/A	N/A

Non-exclusive licence to reproduce thesis and make thesis public

I, Rauno Oja,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

1.2. make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

CALCULATING THE RETURN ON SECURITY INVESTMENT OF RECODING X-ROAD AND ESTONIAN ELECTRONIC IDENTITY SOFTWARE INTO BLOCKCHAIN,

supervised by Oliver Lukason

2. I am aware of the fact that the author retains these rights.

3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, **24th of May 2016**