Kristian Herland

Information security risk assessment of smartphones using Bayesian networks

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 3.8.2015

Thesis supervisor:

Prof. Heikki Hämmäinen

Thesis advisor:

Lic.Sc. (Tech.) Pekka Kekolahti

Aalto University
School of Electrical
Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
MASTER'S THESIS

Author: Kristian Herland

Title: Information security risk assessment of smartphones using Bayesian networks

Date: 3.8.2015 Language: English Number of pages: 8+69

Department of Communications and Networking

Professorship: Network Economics Code: S-38

Supervisor: Prof. Heikki Hämmäinen

Advisor: Lic.Sc. (Tech.) Pekka Kekolahti

The amount of smartphones in use has grown exponentially during the last decade, and with this growth, users have become vulnerable to a new realm of information security threats. Although the security of desktop and laptop computers is relatively common knowledge, smartphone security is not as well understood among end-users, and has not been studied extensively. This thesis comprises an information security risk assessment of smartphone use in Finland using Bayesian networks.

The primary research method in this thesis is a knowledge-based approach to building a causal Bayesian network model of information security risks and consequences. The risks, consequences, probabilities and impacts are identified from domain experts in a 2-stage interview process with 8 experts as well as from existing research and statistics. This information is then used to construct a model which is flexible and lends itself to different use cases such as sensitivity and scenario analysis. This model can also be extended when new data becomes available.

The results show that smartphone use is accompanied by a wide variety of different information security risks such as advanced shoulder surfing techniques and unintentional data disclosure through legitimate applications. Although some risks are difficult or impossible for the user to control, most risks discussed in this thesis are strongly dependent on the user's actions. Therefore, there is a need for increasing security awareness among smartphone users.

Keywords: smartphone, information security, risk assessment, risk analysis, Bayesian network, Bayes.

AALTO-YLIOPISTO
SÄHKÖTEKNIIKAN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Kristian Herland

Työn nimi: Älypuhelinten tietoturvan riskikartoitus käyttäen Bayes-verkkoja

Päivämäärä: 3.8.2015 Kieli: Englanti Sivumäärä: 8+69

Tietoliikennetekniikan laitos

Professuuri: Tietoverkkotalous Koodi: S-38

Työn valvoja: Prof. Heikki Hämmäinen

Työn ohjaaja: TkL Pekka Kekolahti

Käytössä olevien älypuhelinten määrä on kasvanut eksponentiaalisesti viimeisimmän vuosikymmenen aikana ja tämä muutos on tuonut mukanaan täysin uudenlaisia tietoturvariskejä. Vaikka loppukäyttäjillä on keskimäärin kohtuulliset yleistiedot perinteisten tietokoneiden tietoturvasta, ei älypuhelinten tietoturva ole vielä vastaavalla tavalla yleistietoa, eikä sitä ole tutkittu yhtä paljon. Tämä työ koostuu älypuhelinten tietoturvan riskikartoituksesta ja koskee nykyhetken Suomea ja käyttää menetelmänä Bayes-verkkoja.

Ensisijainen tutkimusmenetelmä tässä työssä on kausaalisen Bayes-verkon rakentaminen tutkimustiedon sekä asiantuntijoiden tietämyksen perusteella. Älypuhelinten tietoturvaan liittyvät riskit, seuraukset, todennäköisyydet ja vaikutukset kartoitetaan kahdeksalta aiheasiantuntijalta kaksivaiheisen haastatteluprosessin aikana sekä olemassa olevista tutkimuksista ja tilastoista. Kerätyn tiedon avulla rakennetaan joustava malli, jota voidaan käyttää useaan käyttötarkoitukseen kuten herkkyysanalyysiin ja skenaarioanalyysiin. Mallia voidaan myös laajentaa, kun uutta tietoa tulee saataville.

Työn tulokset osoittavat, että älypuhelimen käyttöön liittyy laaja skaala tietoturvariskejä, kuten edistyneet salakatselun menetelmät sekä tahaton tiedonjako asiallisten mobiilisovellusten kautta. Useimmat tässä työssä käsitellyt riskit ovat vahvasti riippuvaisia käyttäjän omista toimista, vaikka joitain riskejä käyttäjän voi olla vaikeaa tai jopa mahdotonta välttää. Käyttäjien tietoturvatietoisuuden lisäämiselle on näin ollen tarvetta.

Avainsanat: älypuhelin, tietoturva, riskikartoitus, riskianalyysi, Bayes-verkko.

Contents

List of Figures

List of Tables

Symbols and abbreviations

Symbols

s_c	amount of possible states of the child node
s_p	amount of possible states of the parent nodes
n_p	amount of parent nodes

Operators

P(A)	probability of occurrence of event A
P(A|B)	conditional probability of event A's occurrence when event B has occurred

Abbreviations

CVSS	Common Vulnerability Scoring System
CVE	Common Vulnerabilities and Exposures
BN	Bayesian Network
BBN	Bayesian Belief Network
MitM	Man in the Middle
DoS	Denial of Service
IMSI	International Mobile Subscriber Identity
DAG	Directed Acyclic Graph
SSID	Service Set Identifier
WLAN	Wireless Local Area Network
URL	Uniform Resource Identifier
OWASP	Open Web Application Security Project
SMS	Short Message Service
SIM	Subscriber Identity Module
NPT	Node Probability Table
NFC	Near Field Communication
PAN	Personal Area Network
C&C	Command-and-Control
SS7	Signalling System No. 7
OS	Operating System
PIN	Personal Identification Number
LTE	Long-Term Evolution
IDS	Intrusion Detection System
ISO	International Organization for Standardization
IEC	International Electrotechnical Commission
CPU	Central Processing Unit
ENISA	European Union Agency for Network and Information Security
NIST	National Institute of Standards and Technology
OWASP	Open Web Application Security Project
USB	Universal Serial Bus
MDL	Minimum Description Length

1 Introduction

1.1 Motivation

The global number of smartphone users has increased rapidly since the advent of the first Apple iPhone in 2007 and already surpassed 1 billion in 2012 [48]. In Finland, the share of smartphones relative to all mobile handsets in use exceeded 50 % in 2013 [59]. As smartphones have become powerful enough to run most of the typical functionalities used on laptop or desktop computers, users are effectively migrating their computing tasks from traditional computers to smartphones. While traditional computer security is common knowledge and even end-users typically employ security software such as anti-virus on their computer, smartphone security is not as well understood among end-users.

Research concerning specific smartphone vulnerabilities exists in large numbers. Terms such as *mobile malware* and *mobile phishing* already return numerous matches in research paper searches. However, comprehensive risk assessments of smartphone use are not readily available. It is not immediately clear how much mobile malware contributes to the information security breaches that occur via smartphones, for example. Moreover, it is unclear how much smartphone use contributes to all information security breaches.

1.2 Research questions and objectives

The main objective of this thesis is to perform a high-level risk assessment of information security related to smartphone usage. As a secondary objective, this research aims to design and implement a practical risk assessment process for eliciting information from multiple experts and consolidating this information into a Bayesian network. The outcome of this risk assessment is a Bayesian network model of information security risks, which can be used for various purposes such as scenario and sensitivity analysis.

This research aims to answer at least the following questions.

1. What are the most important causes of information security breaches via smartphones?
2. How much do smartphone users' own actions contribute to their information security or lack thereof?

1.3 Scope

The following statements define the scope of risks to be included in this high-level risk assessment:

1. The scope of this thesis includes risks that are directly related to information security, such as network-based attacks or malicious software.

2. The scope of this thesis also includes such risks which are not directly related to information security but which have consequences that are. Examples include technical failures of a device or network.

3. The scope of this thesis excludes all risks that do not fulfil at least one of the two categories defined in statements 1 and 2. Examples include the health risks associated with prolonged smartphone use.

4. Risks that are not relevant to smartphone use in Finland at the time of this research are excluded from the scope of this thesis.

The most important risks identified during this risk assessment are discussed individually in this thesis and in the resulting model. However, other risks also exist which are in the scope of this thesis as defined above. These risks are included in the model as latent risks, which are not necessarily discussed individually but are taken into account when determining the quantitative values of the model. The risk assessment in this thesis is performed at a high level and as such, in-depth analysis of specific technical scenarios or consequences is out of the scope of this thesis.

1.4 Research methods

The research methods used in this thesis are literature review, expert interviews and Bayesian network modelling and analysis. Figure 1 gives a high-level description of the risk analysis process.

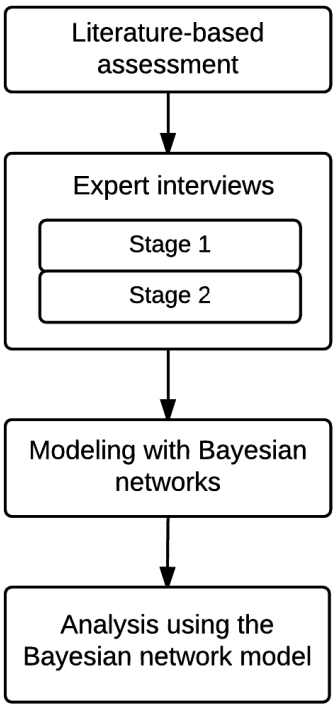


Figure 1: High level overview of risk analysis process.

First, relevant a priori information is reviewed from literature in order to determine the known assets and risks related to smartphone use. This information is then utilized as a basis for interviews with domain experts. Based on the information gathered from the experts, a Bayesian network model of the risks and consequences is created. This model is then used for further analysis of the risks.

The Bayesian network modelling is performed using a software solution named Agenarisk [90]. Alternative software tools for Bayesian network modelling include BayesiaLab [87], Bayes Server [88] and Netica [89], for example. Bayesian networks are chosen as the research method due to their advantages presented in section 2.2.3. Although Bayesian networks have been applied to information security risk analysis, the method has rarely been applied to smartphone security research.

1.5 Thesis structure

The rest of this thesis is structured as follows. Section 2 describes background information concerning smartphones and their usage as well as Bayesian networks and prior research. Section 3 describes the risk assessment process including the literature review, the expert elicitation process and their combined results. Section 4 describes the construction of the Bayesian network model and results derived from the model. Section 5 discusses the applicability, possible use cases and further development of the model and improvements for the process used. Lastly, section 6 concludes and summarizes the thesis.

2 Background

2.1 Smartphone characteristics and usage

The term smartphone is defined in various ways in literature. Some definitions describe these as handheld personal computers [42], some refer to the devices' mobility and some to extensibility with downloadable applications [43]. In this thesis, smartphones are defined as mobile phones that (1) can be extended with third-party-developed applications and that (2) can perform some functionalities usually performed by desktop or laptop computers. The device characteristics of smartphones are discussed in more detail in subsection 2.1.1. Within this thesis, the terms *smartphone* and *mobile device* are used interchangeably.

An integral part of a mobile phone is its operating system (OS), which provides the device's computing infrastructure and defines its application platform. As is discussed in subsection 2.1.2, a significant amount of the users' time on smartphones is spent with downloaded applications. Thus the application platform has a considerable impact on the features, user experience and security of the device.

The three most used mobile operating systems currently are Google's Android, Apple's iOS and Microsoft's Windows Phone [44]. Examples of operating systems with lower market shares include RIM's Blackberry, Nokia's Symbian, Mozilla's Firefox OS and Jolla's Sailfish OS. The momentary distribution different operating systems in use is difficult to estimate, because devices are frequently decommissioned or recycled. However, quite accurate statistics exist concerning the amount of mobile devices sold. Figure 2 shows the distribution of smartphones sold globally in Q4 2014 according to the International Data Corporation [44]. According to figure 2, the majority of devices sold were running Android while iOS accounted for approximately 20 % of devices.

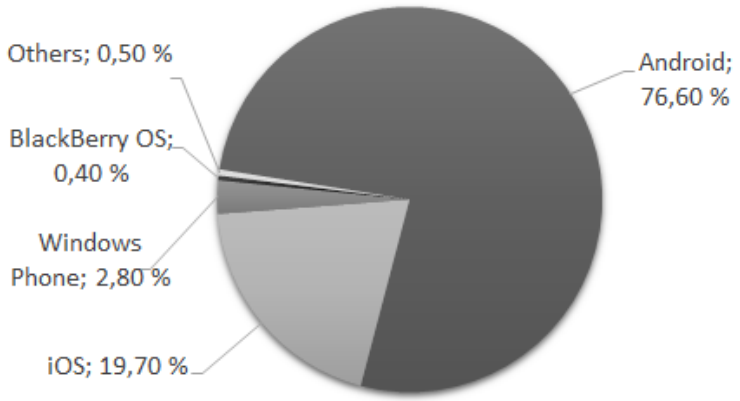


Figure 2: Mobile operating system market share in Q4 2014 [44]

Although mobile applications can be developed by anyone with the required skills, they are usually distributed through a virtual application store managed by the developers of the OS, who have taken different approaches to their application stores. Android has the most open application store [5], the Play Store, in which applications are not subjected to manual verification before initial distribution [45]. Therefore, malicious applications often spread through the application store, although usually for a very limited amount of time [75]. In contrast, Apple's App Store employs a more restrictive approach wherein applications are manually screened pre-emptively [45], which can be very effective at inhibiting malicious applications. However, the insufficient security of an application can be difficult to detect even in manual testing.

2.1.1 Device characteristics

Smartphones typically have a powerful processor with the computing power equivalent to an older desktop computer. The devices are thus able to run a majority of functionalities that are usually performed on desktop or laptop computers. However, the operating system infrastructure also exhibits many differences from traditional computers such as the programming languages supported and the sandboxing of applications that is utilized in most operating systems for both stability and security [46].

Mobile devices also have various built-in security measures such as device passcodes, SIM card PIN codes, application sandboxing, application permissions and storage encryption. A traditional security function is the SIM card PIN code, which prevents many scenarios of unauthorized use of a mobile subscription. Another typical security measure is authentication to unlock the device itself. The authentication method itself can be anything from a numeric code or pattern lock to biometric identification such as a fingerprint.

In order to communicate with other devices and the physical world, modern smartphones support several different communication or data transport technologies. These can be categorized in the following way:

Cellular network interface
: The most traditional cellular technology used by mobile phones for voice calls is GSM. In addition to GSM, smartphones generally support newer cellular technologies with significantly higher bandwidth that are better suited for data transfer. Current technologies available to consumers in Finland such as LTE can support transfer speeds of more than 100 Mbps.

Wireless LAN (WLAN) interface
: Most smartphones include an interface for connections to Wireless Local Area Networks (WLAN), which are based on the standard IEEE 802.11. Several different versions of the standard exist which mainly differ by their transfer speeds. Public WLAN networks are common throughout the world restaurants, hotels and airports, for examples.

Personal Area Network (PAN) interface
Smartphones often support several PAN technologies such as Bluetooth, Infrared or Near Field Communication (NFC). These technologies provide low-cost, ad hoc, short distance connections between devices. Bluetooth is a wireless technology, originally based on standard IEEE 802.15.1, which is widely used for data transfer between a smartphone and other connected devices such as hands-free headsets or smartwatches. NFC is a communication technology used at very short distance, often only a couple of centimetres, and is employed in contactless payment systems.

Each device also has an internal memory whose size usually ranges from several gigabytes upward, which can on some devices be expanded using an external memory card. Due to the extensive storage size, the devices often store large amounts of the user's data as well as temporary data used by the OS and applications. A typical mobile user's device might contain for example all email messages sent and received on his accounts during the previous month.

The primary input interface of modern smartphones is most often a touchscreen. However, devices with various types of keyboards also exist. In addition to the primary input interface, smartphones utilize various sensors for monitoring the physical world. Google's Android's developer documentation categorizes these sensors into the following three broad categories [47]:

Position sensors
Position sensors aim to determine the physical position of a device. The different types of sensors offer different degrees of precision and accuracy, and are often combined in order to obtain reliable results. Examples of specific sensors used for this purpose are magnetometers and orientation sensors. In addition, the Global Positioning System (GPS) and WLAN interface are often also used for determining the geographical position of the device.

Motion sensors
The motion sensors measure real-time movement of the device. This is accomplished by combining measurements of acceleration forces and rotational forces along all three axes. Sensors used for motion sensing include accelerometers, gravity sensors, gyroscopes and rotational sensors [47].

Environmental sensors
Environmental sensors measure various environmental parameters which can include air temperature, pressure and humidity, for example. Other common environmental sensors include illumination sensors and proximity sensors, which aim to determine the devices proximity to other objects such as a table or the user's ear. The microphone and camera of a mobile device can also be interpreted as environmental sensors.

2.1.2 Typical usage

According to global research, 80 % of online users owned a smartphone in 2014 and 75 % of these users accessed the internet through their smartphones monthly [8]. Out of all mobile phones shipped in Q3 2014, 70 % were smartphones [49]. In 2014, a global survey found out that users spend on average 1.85 hours online via a smartphone each day [8].

Due to the mobile nature of smartphones, they are used in various contexts. Table 1 describes five contexts and their respective shares of interaction time with smartphones, as identified by Karikoski and Soikkeli [60] through analysing mobile network cell ID and WLAN data. However, the sample used in the study was biased toward students and thus does not necessarily represent an accurate estimate of usage in the workplace.

Context	**Description**	**Interaction time**
Home	The user's home	53 %
Office	The user's workplace	12 %
Abroad	Not in one's home country	3 %
Other meaningful	A frequently visited context which does not have the characteristics of *home* or *office*	8 %
Elsewhere	Any contexts that do not fit in any of the former categories	24 %

Table 1: Smartphone usage contexts and their respective shares of interaction time [60]

The majority of smartphone users use mobile applications every day [5]. According to research performed in the United States, 89 % of smartphone interaction time is spent using various applications while 11 % of the time is spent on the web [4]. Social networking applications amount to approximately 25 % of the time used on mobile applications while mobile games amount to 16 % [5]. The distribution of users' time on different mobile application categories is further described in figure 3.

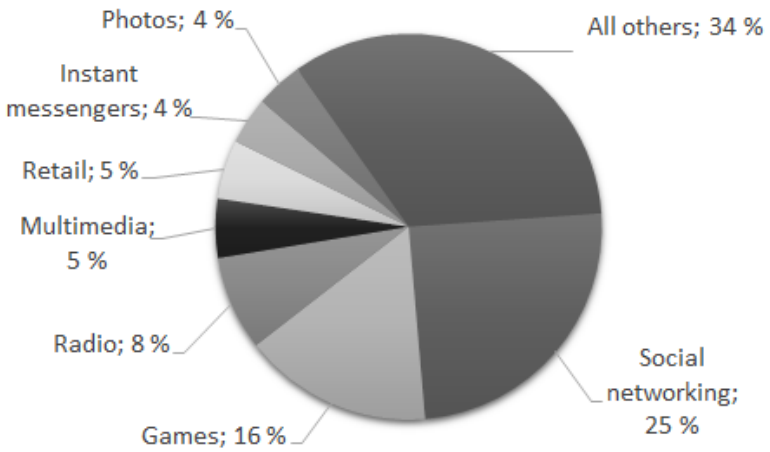


Figure 3: Time spent on mobile applications per application category

Another noteworthy application category consists of the mobile payment systems and banking services whose use has increased rapidly during the previous years. In 2014, 12.4 % of all payment transactions in Europe were already initiated from mobile devices [50] and approximately 50 % of smartphone users in the US had used mobile banking services on their smartphones [7].

2.2 Bayesian networks

2.2.1 Definition

Bayes' theorem is a mathematical application of conditional probabilities, which relates current and prior probabilities. The mathematical statement of Bayes' theorem can be seen in equation 1 below,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \tag{1}$$

where A and B are events. *P(A)* and *P(B)* represent the probability of occurrence for events A and B, respectively, independent of the other event. *P(A|B)* represents the conditional probability of event A's occurrence when given that B has occurred.

A Bayesian network (BN), Bayes network or Bayesian Belief Network (BBN) is a probabilistic model used for portraying variables and their direct dependencies. The model is represented using a directed acyclic graph (DAG) in which variables are portrayed by nodes and their causal relationships by directed edges [22] such as in figure 4. The existence of a directed edge between nodes A and B indicates that the nodes are directly dependent on each other, whereas the direction of the edge indicates the direction of the causal relationship. The edge direction is usually chosen in the direction of cause to effect. In such contexts, node A in figure 4 is referred to as the child node and node B as the parent node. Lack of an edge indicates that there is no direct dependency between the nodes.

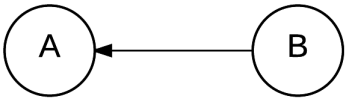


Figure 4: Simple example of Bayesian network

The graphical model in figure 4 does not define any probabilities, states or impacts of the network and is thus referred to as the *qualitative* model of the Bayesian network. To extend this qualitative model, a Node Probability Table (NPT) defines for each node the probability distribution of the node's possible states. If the node has one or more parent nodes, the NPT describes the node's probability distribution conditional on its parents' states. Otherwise, the node's probability distribution is called a marginal probability. The set of NPTs in a network constitutes the *quantitative* model of the Bayesian network.

Bayes' theorem is used as a basis for Bayesian inference, a method of statistical inference, which is used for updating the probability distribution of a variable based on evidence observed [81]. The Bayesian network builds on this inference to create a network of variables that obey Bayes' theorem and are subject to Bayesian inference. A simple example of a Bayesian network is the Naive Bayes classifier, which has been widely used for probabilistic classification.

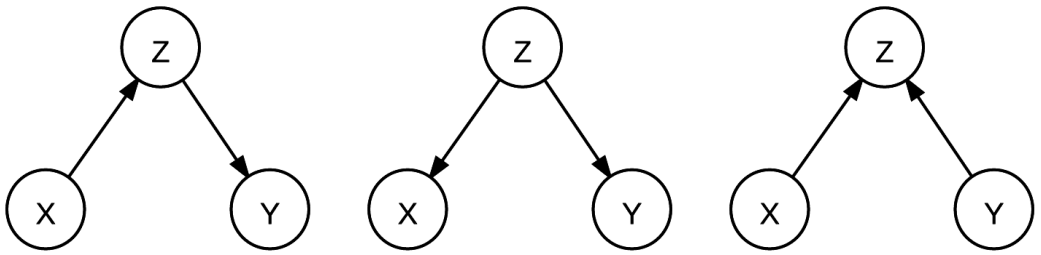
To summarize, the Bayesian network model consists of three parts: (1) the set of variables to be analysed, (2) a graphical structure that describes the dependencies between the variables, hereafter referred to as the *qualitative model*, and (3) a set of conditional probability distributions, one for each variable, hereafter referred to as the *quantitative model*.

2.2.2 Properties and construction

An important attribute of Bayesian networks is the propagation of probabilities between nodes, which contributes to the advantages of using a Bayesian network described in subsection 2.2.3. Probabilities are propagated between nodes with direct connections but also between nodes that are indirectly connected through another node or set of nodes. These indirect connections can be described using a criterion named d-separation [22], which determines whether two sets of nodes X and Y are independent of each other given node set Z.

To illustrate the d-separation of nodes, three essential cases should be discussed. Figure 5a illustrates a serial connection between nodes or node sets X and Y. If the value of Z is known, any knowledge about X is irrelevant to Y, thus X and Y are said to be d-separated given Z. The same statement applies to the diverging connection or common cause case illustrated in figure 5b.

Figure 5c illustrates a converging connection wherein nodes X and Y have a common effect. If no evidence of Z is observed, knowledge about node X is irrelevant to Y. However, if the value of node Z is known, evidence of X affects the probability distribution of node Y. Thus, unlike in the previous two cases, evidence of node X can be transmitted to Y given Z. Nodes X and Y are thus d-connected given Z.



(a) Serial connection (b) Diverging connection (c) Converging connection

Figure 5: Indirect connection scenarios in Bayesian networks

The methods of creating a Bayesian network model can be roughly divided into (1) data-driven and (2) knowledge-based approaches. These methods are sometimes also referred to as bottom-up and top-down approaches, respectively. In the data-driven method, machine learning is used to identify the network structure and parameters. The methods for structural learning can be further divided into two main categories:

(1) CI-test-based approaches and (2) optimization-based search methods [84]. The first method uses conditional independence (CI) assumptions to infer the structure of the network [84], while the second uses a search strategy and scoring function to identify the network structure that best fits the given set of data [85]. The scoring functions often utilize the minimum description length (MDL) principle introduced by Jorma Rissanen [91]. Parameters are then estimated from data using methods such as the maximum likelihood approach [92].

In the knowledge-based approach, the network structure is determined using causal knowledge of domain experts [86]. The parameters are identified from existing data or elicited from domain experts, often combining both sources. These methods are further discussed in section 3.3.

2.2.3 Advantages and disadvantages

Classical methods of causal and frequency analysis include, in addition to Bayesian networks, Fault Trees [78], Markov chains [79] and Petri nets [80]. Fault trees are commonly used in causal analysis but can be completely replaced by Bayesian networks. Markov chains and Petri nets on the other hand are more frequential methods and are not suitable for causal analysis.

Compared to alternative methods, Bayesian networks exhibit the following advantages:

1. Efficient consolidation of hard data and expert opinion [82].
2. Ability to capture causal knowledge even from domain experts with little or no statistical experience [23].
3. Easily understandable format for visualizing causal relationships between variables [22].
4. Suitability for simple expert elicitation methods [24].
5. Robustness with regards to incomplete information [82].
6. Flexibility and abundance of use cases [24].
7. Support for structural learning [82].

However, Bayesian networks also exhibit the following disadvantages:

1. Continuous variables must be discretized before use [51]. However, Bayesian networks are mainly a classification method, wherein this disadvantage is negligible.
2. Determining quantitative values via expert elicitation is a complex [82] and time-consuming process [22]. However, the number of variables to be elicited can be reduced using methods discussed in section 3.3.3.

2.2.4 Prior research in risk management

According to a meta-analysis of Bayesian networks research performed by Weber et al. in 2012 [24], the number of literature references to Bayesian networks applications in risk analysis have risen considerably during the last decade. Some of the first documented applications of BNs in risk analysis were a decision support system to evaluate terrorism threats by Hudson et al. in 2001 [26] and risk assessment of structures under fire by Gulvanessian and Holicky in 2002 [25].

After the first applications during the beginning of the century, BNs have been utilized in many fields for risk analysis. In 2004, Cornalba and Giudici [28] described a knowledge-based Bayesian networks method which combined statistical data with expert opinion in order to determine the risks related to banking organizations. In 2006, Kim and Seong [27] proposed a method for assessing the safety of nuclear power plants in various contexts with a knowledge-based Bayesian network approach. According to the authors, a significant advantage of this method was the ability to take into account the effect of interdependency between critical systems and human operators.

Bayraktarli et al. [52] employ Bayes networks for risk management related to earthquakes whereas Straub [53] describes an application of Bayesian networks to risk assessment of various kinds of natural hazards. Cheon et al. [34] propose a method for predicting high concentrations of ozone using Bayesian networks. This method could also be employed for other rare event prediction purposes such as fraud detection or diagnosis. Hanea and Ale [33] discuss an approach to estimating risks of building fires in Netherlands using Bayes networks.

Bayesian networks have also been used in the maritime field. Russell et al. [29] describe a methodology for applying Bayesian networks to assessing the reliability of search and rescue operations based on expert judgement while Trucco et al. [30] use expert elicitation to construct a quantitative Bayesian network describing the risks of maritime transportation. Eunchang et al. [54] describe a significant application of Bayes networks to risk management of a large engineering project. The application to the risk management of the Korean shipbuilding industry included surveying 252 industry experts. Bayesian network models have also been used for assessing the probability of ship collision [94] and effectiveness of oil combating [95] in the Gulf of Finland. Both studies combine prior statistical data and expert opinion to create a quantitative safety assessment.

Duijm [31] describes the qualitative use of Bayesian networks in safety-barrier diagrams, which are easy-to-understand graphical models used in risk analysis and safety management. Røed et al. [32] discuss a framework called hybrid causal logic, which combines traditional risk analysis tools with causal knowledge-based Bayesian networks. They also review an application of the framework to both the aviation industry and the offshore oil and gas industry.

Peltola and Kekolahti [76] use Bayesian networks for risk assessment of the Finnish TETRA PSS network, where 10 risk sources such as sabotage contribute to unavailability of the network. The authors use expert knowledge to construct a quantitative BN model, which is used to analyse the effects of different risk controls such as access control on service availability.

2.2.5 Prior research in information security

A paper by Mo et al. [35] proposes a quantitative model for evaluating a firm's cyber security readiness by use of Bayesian networks. The proposed model uses a firm's security profile and data breach statistics as input to a Bayesian network model of vulnerabilities, threats and risks. The model gives as output a universal risk score that could be used to evaluate the state of information security in firms.

A topic that has been widely discussed in research papers during the previous five years is modelling the interdependence of vulnerabilities in IT with Bayesian networks. Although individual vulnerabilities can be exploited for limited gain, combining vulnerabilities can allow an attacker to penetrate several layers of defence for example in an enterprise network. These attack scenarios can be described using attack graphs.

Figure 6 describes an example network structure (left) and a respective attack graph (right). In this example, exploiting either vulnerability 1 or both vulnerability 2 and 3 on the web server allows access to the file server, where vulnerability 4 can be exploited. This in turn allows the attacker to exploit either vulnerability 5 or 6 on the end-user devices in the internal network.

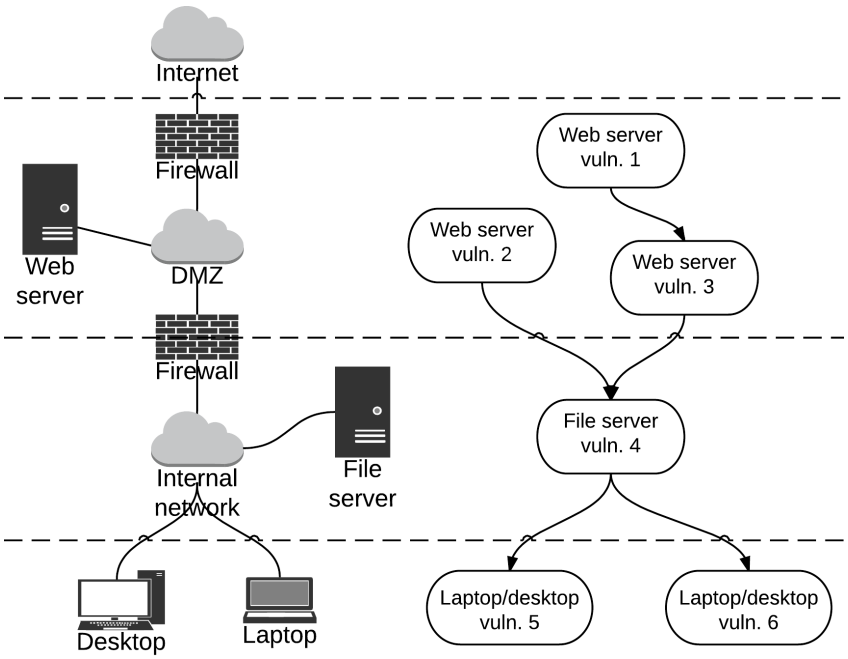


Figure 6: Example attack graph

An attack graph such as figure 6, or a set of attack graphs, can be further extended into a complete Bayesian network by determining the marginal and conditional probabilities of the nodes such as done by Xie et al. [41]. Most known vulnerabilities have been studied and evaluated according to standards such as the Common Vulnerability Scoring System (CVSS) which is used by several public Common Vulnerabilities and Exposures (CVE) databases to represent metrics such as exploitability and impact. These databases can be leveraged for populating a Bayesian network's NPTs automatically. [38]

Dantu and Kolan [39] extend the Bayesian network model of attack scenarios with different attacker types such as a corporate insider or a political "hacktivist". Noel et al. [36] use the Bayes network model to evaluate expected return on investment regarding different information security investments. Farmad et al. [37] also describe a quantitative method for performing cost-benefit analysis as part of the Bayes network vulnerability assessment.

The BN models of attack scenarios could also be utilized in intrusion detection systems (IDS) for determining the most likely attack path in the event of a breach. Xie et al. discuss the possibility of using this information for mitigating attacks in real-time as effectively as possible [41].

Sommestad et al. [40] present a framework for analysing cyber security using Bayesian statistics. The proposed method merges Bayesian attack networks with abstract models to describe a system's cyber security. Wang and Guo [77] propose a method of using Bayes Networks to automatically categorize software vulnerabilities based on their type.

This high-level risk assessment does not include analysis of specific technical attack scenarios and individual vulnerabilities.

3 Risk assessment

3.1 Risk assessment approach

Risk assessment is usually defined as a method for identifying the threats that a specific group of assets are vulnerable to and determining the threats' probabilities and impacts on each asset. On the other hand, the ISO/IEC 27005 risk management standard refers to this aforementioned process as risk *analysis*. According to this widely used standard, risk *assessment* additionally includes evaluating these risks against the risk evaluation and acceptance criteria. The standard also defines a task named risk *treatment* which includes evaluation of actions that could help avoid, reduce or transfer the risk.

Peltier [9] defines the information security risk assessment process as the following 6 steps. Figure 7 describes these steps in the form of a Bayesian network.

1. **Asset definition** is the step where assets are identified and defined. An asset in this context could be the availability of the smartphone.
2. **Threat identification** includes identifying all undesirable events that affect the one or more assets in a negative way. A threat in this context could be losing the smartphone.
3. **Determining probability of occurrence** is the action of assigning a probability to each threat defined in step 2. These probabilities can be conditional on several different variables.
4. **Determining the impact of the threat** is the process of determining which assets each threat affects and with what severity.
5. **Controls recommended** is the step where mitigating controls and safeguards are identified. A control in this context could be to not leave the smartphone unattended.
6. **Documentation** of the risk assessment results.

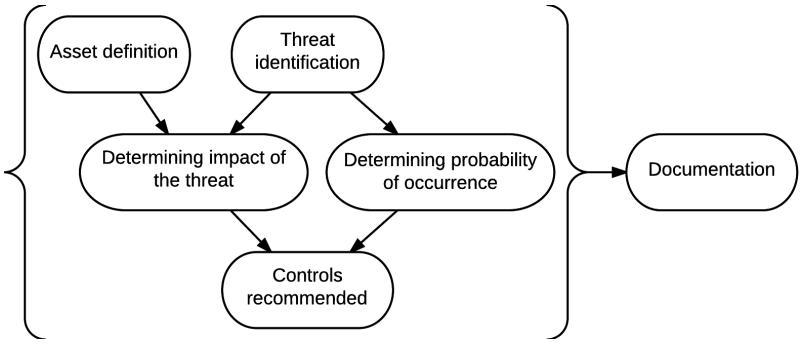


Figure 7: Information security risk assessment process as defined by Peltier [9]

3.2 Literature-based assessment

3.2.1 Assets

In risk analysis, assets are subject to risks. If a risk is realized, an asset is affected in some negative way. Thus by definition, assets are something that should be protected from risks, which can be accomplished by reducing the probability or impact of the risk. In addition to assets, this thesis also discusses consequences, which can be viewed as risks' effects on assets.

A classic categorization of the general information security goals or attributes is the triad *confidentiality*, *integrity* and *availability*, often dubbed "CIA". These attributes are closely related to the definition of assets as they are what is generally required of assets.

Many different categorizations of smartphone assets can be found in literature. One example of the simplest categorizations is the one used by Jeon et al. [17] below:

1. Private information
2. Device
3. Applications

In this approach, one category includes all information that should not be made public, which can be of various forms and sources. Some of the information is deliberately stored on the device or SIM card by the user such as contact information, sent and received SMS messages as well as documents stored on the device. Other types of data can be less visible to the user such as temporary files, browser cache or data stored by a persistent login functionality. Also, some data is not stored on the device at all and is only available in real time such as usage information and location data.

Another category in the list by Jeon et al. is defined for the smartphone device itself. The smartphone device can evidently have both financial value to the user as well as be valued for its availability. The availability aspect of a smartphone not only comprises the physical availability of the device but also the availability of its functionalities to the user. Thus the authors also include in this category system resources such as the battery charge, CPU computing power and memory, which are required for the smartphone to fulfil its duties to the user.

The third and last category defined by Jeon et al. comprises applications. The authors reason that applications should be included as an individual category because of two reasons; (1) commercial applications can have a price and (2) applications can store and handle sensitive information.

In a security assessment related specifically to Android devices, Shabtai et al. [13] define a similar list to that of Jeon et al. In this list, which can be seen below, the device hardware and device resources are defined as separate categories:

1. Private/confidential content stored on the device
2. Applications and services
3. Device resources (battery power, communication, processing power etc.)
4. Hardware

The categorization by Shabtai et al. includes one broad category that comprises all applications and services, i.e. virtually all functions that can be performed by users on smartphones. While the categorization by Jeon et al. focused on the data and cost related to applications, Shabtai et al. also view the availability of applications and services as an asset.

Theoharidou et al. [10] propose a personal risk assessment method for smartphones and therein define the following four categories of assets:

1. Device
2. Connectivity
3. Data
4. Applications

The main difference between this and the previous two categorizations is that the authors define connectivity as an individual category. A wide variety of functionalities make use of the connectivity methods described in section 2.1.1 and as such, the availability of the connections can be viewed as an asset. On the other hand, the confidentiality and integrity of data transfers are also of value to the user. A user would for example prefer that instant messages reach the recipient confidentially and unaltered.

Similarly to the approach of Theoharidou et al., Ledermüller and Clarke [20] separate connectivity channels from the other assets. However, Ledermüller and Clarke divide this category further into (1) communication by voice and messaging and (2) network data access. According to the authors, the following asset categories were chosen based on the identified mobile phone usage trends and application market offering:

1. Communication by voice and messaging
2. Network data access
3. Applications
4. Device and stored data

A report by the European Union Agency for Network and Information Security (ENISA) [11] defines the following categorizations of smartphone assets:

1. Personal data
2. Corporate intellectual property
3. Classified information (governmental)
4. Financial assets

5. Device and service availability and functionality
6. Personal and political reputation

Compared to the other categorizations presented, the categorization by ENISA focuses less on the technical aspects of a smartphone and more on the contexts that smartphones are used in. This list introduces three separate categories of data, based not on the type or form of the data but rather the source and impact of the data. Most data stored on smartphones likely relates either to the user's personal life or work. The separation between the sources of data demonstrates the different value and degree of confidentiality assigned to and expect from data from different sources. Also, business or governmental data stored on a smartphone is more likely to also be backed up elsewhere than personal data. Thus, the value of the integrity of data depends on the source. The categorization used by ENISA even separates corporate data from governmental data due to the different implications of a confidentiality breach of the data.

In addition to the assets technically associated with a smartphone, risks related to smartphones can also have direct financial consequences. The device itself has a value and will most likely have to be replaced if it is stolen. Also, the user's financial assets are related to the smartphone through mobile subscription billing and possible mobile payment systems. As the final category of the list, ENISA introduces personal and political reputation. The reputation of a user could be affected through the disclosure of personal information as well as impersonating a user through their smartphone.

For the purpose of this risk assessment, the following list of 6 assets was defined:

1. Confidentiality of personal data
2. Confidentiality of business or governmental data
3. Integrity and availability of data
4. Financial assets
5. Availability of device and functionalities
6. Access to other devices and services

According to this categorization, the *confidentiality* of personal data and business or governmental data are separated as two different assets. Reasons for this include the different implications of a confidentiality breach of the data based on its type as well as the different forms that the data exists in depending on its source. Personal data can include for example photos, browser history and sensor information such as location data, which can be collected from the device over a period of time. On the other hand, confidential data is mostly comprised of emails and documents stored on the device. Both categories are likely to include data transmitted over communication interfaces.

A third category comprises the *integrity* and *availability* of all different types of data. In the context of this assessment, this asset mostly concerns data stored on the

actual device. Separate categories are also introduced for financial assets as well as the availability of the device and its functionalities. Financial assets can be affected by events such as excessive charging through the user's mobile subscription, extra charges in a mobile payment system and repair or replacement costs of the device. The availability of the device and its functionalities can be completely lost due to theft of the device or partly diminished due to a technical failure, for example.

In addition to the five first assets, smartphones are also used for accessing a plethora of other services used for various purposes. Possible services on smartphones include voice calls, SMS messages, video streaming services, social networks and online banking, for example. Many applications or services assume that authentication of the user is performed on device level and thus no additional authentication is required before allowing the user to access applications or functionalities. A user might also store usernames and passwords on their device, giving an attacker access to also those services. The implications of a security breach concerning these services depend completely on the nature of the services enabled on the device and the security measures employed in these. Thus a sixth asset category is introduced for the security of access to other services and devices.

In addition to naturally being dependent on the risks, consequences are in some cases also dependent on each other. For example the leakage of credit card information can result in financial consequences. However, the consequence categories were defined so that these dependencies would be minimized. An in-depth evaluation of these dependencies is not performed in this thesis.

To summarize, the following 6 smartphone assets were identified: (1) confidentiality of personal data, (2) confidentiality of business or governmental data, (3) integrity and availability of data, (4) financial assets, (5) availability of device and functionalities and (6) access to other devices and services.

3.2.2 Threats

Information security threats are events that have an undesirable effect on one or more assets, a probability of occurrence and an impact. The probability of a threat can be conditional on many other variables such as whether another risk has been realized or not. The impact of a threat can vary based on the asset examined.

The threat space of smartphones is quite diverse due to their mobility, high computing power and versatile purposes of use as described in section 2.1. A guideline for managing the security of mobile devices by the National Institute of Standards and Technology (NIST) [21] raises many important high-level threats. Mobile devices usually lack physical security controls that have been characteristic of other IT devices and in addition to being used at the office and at home, a smartphone is often used in various public places and means of transportation, where the devices can be connected to untrusted networks. The use cases of smartphones include a large

amount of potentially untrusted applications and interaction with other untrusted systems. Also, smartphones can encounter untrusted content of new types such as Quick Response codes (QR). Another new aspect of the security of smartphones are the use of location services.

Threats can be categorized in several ways such as based on the assets they impact, the type of impact that they have or the realm in which they exist. Theoharidou et al. [10] divide threats into the following four realms and for the sake of readability, the threats are discussed roughly in this order:

1. Device
2. Network connectivity
3. Operating system
4. Applications

The smartphone device itself is vulnerable to many physical threats such as loss or theft of the device. Accurate statistics of smartphone thefts and losses is scarcely available, however, according to survey research published by Consumer Reports [62], 3.1 million smartphones were stolen and 1.4 million lost in the US during 2013. According to Statista [61], the total number of smartphone users in the US in 2013 was 144.5 million and thus approximately 3.1 % of smartphones in total were lost or stolen. According to a benchmark study by McAfee and Ponemon Institute in 2011 [69], employer-provided smartphones account for 62 % of all smartphones in use and approximately 4.3 % of these are lost or stolen each year. This supports the above approximation of 3.1 %.

Another type of physical threat is unauthorized physical access to the device, which can happen for example when the device is stolen or lost and found by another person. If a device does not require authentication such as a security code or facial recognition before usage, an unauthorized user could access the device immediately after gaining physical access. Symantec performed in 2012 an experiment [83] where 50 smartphones were intentionally lost and then remotely monitored for activity. The study found that 89 % of these devices were accessed for personal data and 83 % for corporate-related data by the finders.

A report by the European Union Agency for Network and Information Security [11] discusses improper decommissioning of a device, which refers to the disposal of the phone in such a way that another party is able to recover information from the device. The confidentiality of data can thus be compromised even after a device is decommissioned or recycled.

Subsection 2.1.1 briefly discussed the large amount of sensors integrated into modern smartphones, which can collect an immense amount of data about the user. Even if this data is never stored permanently on the device, an attacker with real time access to the device, for example through surveillance software, can collect this information.

Surveillance software can be installed either by gaining physical access to the device or through malware distribution method. Many legitimate applications exist that can be misused for covert surveillance.

Wang et al. [18] also introduce battery exhaustion attacks as a threat. In order for a smartphone to fulfil its duties to the user, a smartphone requires among other things battery power, computing resources and free memory. Attacks that drain or otherwise overuse these resources must also be considered.

While traditional computers are relatively rarely used in public places or transportation, smartphones can and often are used anywhere and everywhere. This raises a concern of other people eavesdropping on phone calls and watching a user use his smartphone, i.e. "shoulder surfing". Business users who use their laptops on airplanes etc. often protect their screen with a privacy filter that makes it more difficult to view the content of the screen from the sides. However, such privacy filters have not become common for smartphones.

Although it might be difficult to accurately spot the username and password written by a fellow bus passenger with a naked eye, shoulder surfing can also be performed in more advanced ways. A team of researchers from the University of North Carolina developed a mechanism [70] for detecting passwords written using cheap mobile video cameras. In addition to the method being surprisingly accurate, the team was even able to detect passwords by recording a reflection of the smartphone screen from a user's sunglasses.

Smartphones use several different connectivity channels for data transfers such as described in section 2.1.1, and each packet sent on these networks is vulnerable to attacks. On unencrypted WLAN networks, such as the public networks found at restaurants and airports, sniffing other user's data transfers is feasible even without special skills or equipment. Such scenarios could also lead to a Man in the Middle (MitM) attack. Due to the reliance of smartphones on their connectivity channels, smartphones are also vulnerable to network-based Denial of Service (DoS) attacks as well as situations where a network is congested for other reasons. A recent example of a DoS vulnerability is the iPhone notification centre bug wherein most iPhone devices could be crashed by sending a text message that includes certain Arabic characters. However, a more sophisticated attack could do much more than crash individual smartphones.

A risk assessment by ENISA [11] raises another concern related to the connectivity channels and mobility of a smartphone. By spoofing an access point that a user connects to, an attacker will be able to read and intercept all unencrypted data sent on the channel. In the case of WLAN access points, an attacker could simply choose an SSID similar or identical to that of a trusted network. Incautious and security unaware users would likely choose this network for internet access. On the other hand, it can be possible to spoof an access point or base station in

such a way that the smartphone itself cannot distinguish between a legitimate and illegitimate access point. A recent example of such an attack is the secret surveillance uncovered by Aftenposten in Norway in 2014 [56]. Their tracking revealed several IMSI-catchers, which are essentially fake base stations used for data collection.

Several threats also specifically concern the cellular network connectivity of a smartphone. Old versions of SIM cards have been particularly vulnerable to cloning, wherein an attacker could impersonate another user on the network. Milligan and Hutcheson [14] as well as Wang et al. [18] raise a concern that given access to the right equipment, an attacker could eavesdrop signals on the cellular network in order to listen in on GSM voice calls.

Wireless connections on the GSM network between smartphones and base stations are encrypted using stream ciphers such as A5/1, A5/2 and A5/3. Several vulnerabilities exist in all of these such as the ciphertext-only attack documented by Barkan et al. [74]. Unlike most other known methods for compromising GSM encryption, the ciphertext-only attack is feasible in real-time and does not require significant resources. Compromise of the GSM encryption enables various attack scenarios such as call hijacking, altering of data messages and call theft [74].

Several specific risk assessments have been performed regarding the risks related to individual mobile platforms. Android is likely the most researched platform due to its wide usage as well as its relatively open application development and distribution policy. Examples of the Android risk assessments can be found in [13] and [19] and an analysis of Android smartphone security mechanisms in a paper written by Khan et al. [15]. A comparative security evaluation of the most common mobile operating systems with regard to malware has been performed by Mylonas et al. in 2011 [16]. The research included a case study wherein average developers were tasked with implementing and distributing location tracking malware on the different platforms. The evaluation found that the only security measures that prevented the successful implementation and distribution were Apple's iOS's installation requirements and manual pre-distribution-testing.

One threat that users might not be aware of is the unintentional disclosure of data through legitimate applications [11]. Users may not be aware of the privacy policy or privacy settings of applications and thus unwillingly and unknowingly share information with the application developers or third parties. It is common for users to install applications without first reading the respective privacy policy or terms and conditions. The use of cloud services also introduces new risks as all data that is transferred to a cloud service, whether unbeknown to the user or not, is vulnerable various security risks that affect the cloud service and its infrastructure.

Phishing is a threat that has long existed on computers and is also discussed in many smartphone risk assessments [10][11][17][18]. The relatively small screens of smartphones could facilitate these phishing attacks as the user is not necessarily

shown all information necessary for identifying a phishing attack. For example, a browser application might hide the URL and certificate information, or an email application might hide the sender's address and instead only present the sender's name. According to research by Trusteer [65], mobile users who access a phishing site are three times more likely to submit private information than desktop users.

According to research by IBM [66], approximately 50 % of the private information harvested during a single phishing campaign, is submitted during the first hour. After the first hour, phishing sites are often already identified by IT security vendors. The continuous real-time use of smartphones makes it more likely that a phishing campaign reaches the user within an hour on their mobile device than on a traditional computer.

The non-profit OWASP organization maintains a yearly updated list of the most important technical risks that affect mobile applications. The top 10 mobile application risks from year 2014 include the following risks [12]:

1. Weak Server Side Controls
2. Insecure Data Storage
3. Insufficient Transport Layer Protection
4. Unintended Data Leakage
5. Poor Authorization and Authentication
6. Broken Cryptography
7. Client Side Injection
8. Security Decisions Via Untrusted Inputs
9. Improper Session Handling
10. Lack of Binary Protections

A significant amount of threats listed by OWASP have a root cause in insecure coding methods. Mobile applications can be produced by anyone with basic coding skills and are not necessarily suspect to thorough security testing. Even some of the most popular mobile applications have been found to have critical security flaws [13].

In addition to insecurely designed legitimate applications, applications can also be designed to be purposefully malicious. According to a report by Alcatel-Lucent [63], 0.68 % of all smartphones were infected with malware at the end of year 2014. However, the annual Verizon Data Breach Investigation Report [64] found that only 0.03 % of devices were infected with truly malicious code whereas the rest of the infections were less aggressive pieces of software such as spyware or adware. Mobile malware also varies based on its spreading method. The three most common types are described in table 2.

	Disguised	**Covert**
Standalone	Standalone application which has or is claimed to have a useful functionality which causes users to install it willingly	Covert application which is installed without user consent
Embedded	Malicious code is inserted into a legitimate application which is then repackaged and redistributed	N/A

Table 2: Mobile malware spreading methods.

The redistribution of legitimate software with malicious code embedded would likely be identified quickly in official application stores. However, users also install applications from outside the official stores. Most Android users are able to install applications from outside their application store even without removing device security restrictions.

Malicious software on smartphones can have several purposes. A report by ENISA [11] refers to three distinct kinds of malware: (1) spyware, which is meant to stealthily collect data from a large amount of users, (2) diallerware, which causes billing by calls and/or messages to premium services and (3) other kinds of financial malware such as those that collect credit card information when users make purchases on their smartphones. Becher et al. [58] also make a distinction between diallerware and other financial malware, which could also include attacks targeting mobile payment systems such as Apple Pay or Danske Mobile Pay. Malicious applications can also have other purposes such as corrupting data on the device, causing technical failures or making the device part of a botnet.

Mobile malware has also been found to target mobile banking services. In 2014, a mobile malware application targeting South Korean mobile bank users spread. This application provided the user with a screen quite similar to that of a South Korean bank's legitimate application. However, when the user entered their credentials, these were sent to a C&C maintained by the attackers [71].

The possible consequences of malware and vulnerabilities in legitimate software can be dramatically increased if security and privilege restrictions imposed on the device have been removed, which is usually referred to as "rooting" or "jail breaking" the device. For example Android devices run each application in a dedicated Dalvik virtual machine where it only has access to its own application data. However, this restriction can be removed with the right skills and equipment.

After careful review of the discussions in existing literature, the following 19 information security risks related to smartphone use were identified:

1. Attacks on decommissioned devices
2. Client side code injection
3. Cloning SIM card
4. Denial of Service
5. Diallerware
6. Eavesdropping
7. Loss or theft of device
8. Malware (excl. spyware, diallerware)
9. Mobile payment systems abuse
10. Network spoofing attack
11. Phishing
12. Resource abuse
13. Shoulder surfing
14. Sniffing
15. Spyware
16. Surveillance
17. Technical failure of device
18. Unauthorized physical device access
19. Unintentional data disclosure

In addition to the risks discussed above, smartphone users can also be vulnerable to many other risks that are not related to information security such as the health effects of smartphone use. However, the risks that are unrelated to information security, are out of the scope of this thesis as defined in section 1.3.

3.3 Expert elicitation

The purpose of the expert interviews was to collect the data required in order to construct a Bayesian network model of information security risks and assets related to smartphone use in Finland at the time of this research. To accomplish this task, experts were queried about both the qualitative structure of the network and the quantitative information i.e. the probabilities, impacts and strengths of mutual dependencies of the risks as defined in subsection 2.2.

The interview process was divided into two stages. The purpose of the first stage was to gather enough information to build a qualitative model of the information security risks and consequences, i.e. the graphical BN structure in which nodes represent risks or consequences, and edges indicate causal relationships. During the second stage, this model was presented and validated with each expert after which the strengths of dependencies and impacts were determined. In addition, both stages of the interview process were first tested with a non-expert. Figure 8 shows a visualization of this process.

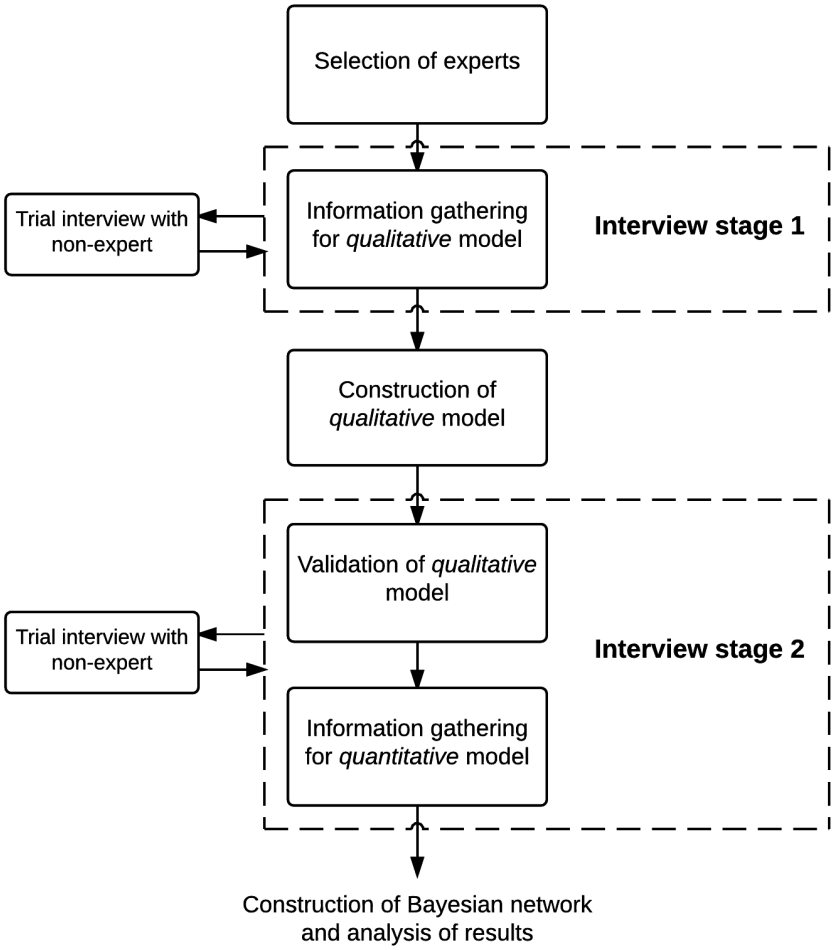


Figure 8: High level overview of expert interview process

A list of common risks and assets related to smartphone use was created based on the literature assessment performed in subsection 3.2 and used to categorize the risks and assets brought up by the interviewees. However, in trial interviews with non-experts it was found that the term asset was difficult to define and understand unambiguously in this context. Thus the assets were translated to consequences of risks, which seemed easier to understand and discuss. The translations from asset in subsection 3.2.1 to consequence are listed in table 3.

Asset	**Consequence**
Confidentiality of personal data	Leakage of personal data
Confidentiality of business or governmental data	Leakage of confidential data (business/government)
Integrity and availability of data	Data loss or corruption
Financial assets	Financial consequences
Availability of device and functionalities	Unavailability of device or services
Access to other devices and services	Unauthorized access to other devices or services

Table 3: Mapping between risk consequence categories and asset categories

3.3.1 Experts

The experts to be interviewed were chosen based on experience, subject matter knowledge, current employer and position. Two main objectives were taken into account when choosing the experts: (1) the group of experts should include various specializations with partial overlap in order to ensure completeness of the information available and (2) the interviewee's experience and knowledge in the domain and their own viewpoint should be good or very good in order to ensure the quality of the information available. The list of experts is described in table 4. In this thesis, each expert represents only their own subject matter knowledge and not in any way their employer.

#	**Title, Organization**
1	Manager, Deloitte
2	Professor, Aalto University
3	Manager, Deloitte
4	Information Security Officer, RAY
5	Senior Manager, Product Security & Privacy, Microsoft
6	Senior Researcher, Aalto University
7	Information Security Specialist, National Cyber Security Centre of Finland (NCSC-FI)
8	IT Security Manager, TeliaSonera Finland

Table 4: Interviewed experts

3.3.2 Interview stage 1

The purpose of the first stage interviews was to elicit from experts the information required for building the qualitative causal model of the Bayesian network described in subsection 2.2, i.e. a list of the variables involved as well as their causal dependencies. In order to capture this information from expert interviews, two different types of techniques are commonly used. *Structured* techniques involve specific questions about predefined concepts and are thus most suitable for confirming existing knowledge. *Unstructured* approaches on the other hand focus on exploring new information and are thus well suited for use in domains for which existing knowledge is lacking or non-existent. Nadkarni and Shenoy illustrate the unstructured approach with the following example question: "What are the factors relevant to the decision?" [23]

Prior knowledge about the domain of this risk assessment was available but in limited extent. Information security risks related to computing have been studied extensively and research also exists concerning the security of smartphones. However, the individual research papers concerning smartphone security only discuss fragments of the whole risk space. Due to the availability of prior information but lack of

completeness, a combination of *structured* and *unstructured* methods was used. The interviewees were asked open questions about which threats exist in the smartphone risk space in today's Finland and what consequences relate to these. If required, the interviewees were also shown the predefined list of threats described in subsection 3.2.2.

The first stage interviews roughly followed the process visualized in figure 9 and described below. However, due to time limitations and varying amount of input available from experts, some interviews followed this process only superficially. Regardless, as much information as possible was gathered concerning each phase with each expert. Before interviews, the experts were given prior information about the topics to be discussed in the interview.

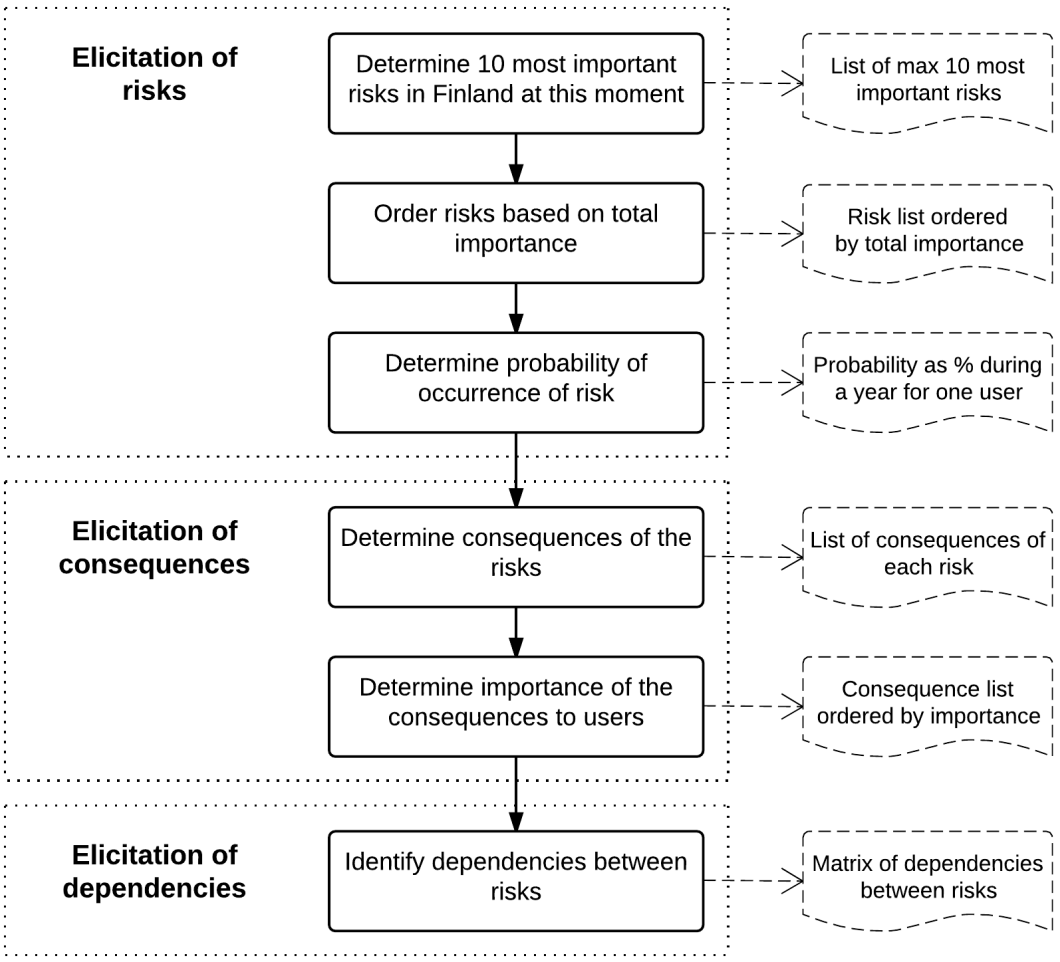


Figure 9: Expert elicitation stage 1 interview process

During the interviews, each expert was asked to list, according to his or her own expertise, the 10 most important information security risks related to smartphone use in Finland at the moment. Risk importance was defined to be based on both probability and impact, i.e. total importance, and risks were discussed to ensure

mutual understanding. In most interviews, the list of risks defined in subsection 3.2.2 was reviewed in order to evaluate its completeness and map the discussed risks to the risks in this list. The interviewees were also asked to order the risks based on total importance.

Next, the interviewees were asked to approximate how many times each risk can be expected to occur during a year in a random sample of 10000 smartphone users in Finland at the moment. However, determining a realistic estimation of the exact volume of occurrence of a risk event is difficult even to domain experts in the absence of statistics. The interviewees were then asked to list the possible consequences of each risk without considering the severity of the consequences and to identify whether the risks have any consequences that were not included in the consequence categories defined in table 3. The interviewees were also asked to order the consequences based on average importance to users. Lastly, each interviewee was asked to identify all dependencies between risks which have a non-negligible strength.

One of the challenges of the interview process was how to clearly define the objectives of the interview without introducing bias. An important part of the interview was the separation between threats or risks and consequences and thus the use of an illustrative example was considered. However, while testing the interview process on non-experts, these examples were observed to introduce anchoring [93], and thus possibly bias, regardless of whether the example related to information security or not. For this reason, an illustrative example was not used.

An Excel tool was designed to facilitate the interview and is described in appendix B. Using this tool, the information needed in latter steps was automatically populated after completion of the prior steps. For example the matrices used for gathering information on conditional relationships were automatically filled in with the risks and assets listed by the interviewee.

3.3.3 Interview stage 2

The main objective of the second stage interviews was to collect the information necessary in order to construct the quantitative Bayesian network model. The quantitative model, as defined in section 2.2, is the set of Node Probability Tables (NPT) assigned to the nodes of the network structure and the necessary information thus consists of the probabilities, impact strengths and strengths of dependencies between risks in the network. In this model, each risk was defined as a Boolean node and each consequence as a ranked node.

The impact of a risk is often measured on scales such as *low-medium-high* or *very low-low-medium-high-very high*. A scale with an odd number of steps is usually preferred due to the advantage of having a middle choice, which generally makes the scale easier to use. According to Fenton et al., experts are rarely satisfied with 3-point scales [55], however, a 5-point scale increases the difficulty of choosing an

impact strength. A typical 3-point scale was chosen due to its simplicity and unambiguousness. In addition to the *low-medium-high* scale, some risks examined in this assessment do not necessarily result in an impact when realized. Thus a fourth step was added to the scale to describe situations where the impact is non-existent or negligible. The impact scale of each consequence node was defined as shown in table 5.

Scale	Impact
Negligible	Negligible or non-existent impact
Low	Minor impact
Medium	Notable impact
High	Substantial impact

Table 5: Consequence severity scale

Several different methods exist for eliciting the content of NPTs. Manual elicitation by interview is possible but quickly becomes infeasible in non-trivial networks due to the exponential increase in NPT size with the amount of nodes. The size of a node's NPT follows equation 2 below,

$$size = s_c * {s_p}^{n_p}, \tag{2}$$

where *size* denotes the amount of values in the NPT, s_c the amount of possible states of the child node, s_p the amount of possible states of the parent nodes and n_p the amount of parent nodes. The equation assumes that all parent nodes have the same amount of possible states.

The interviewed domain experts are likely busy and a very time consuming process can thus not be used. Also, the subject matter experts are not necessarily experienced in statistics and probability theory. Thus it makes sense to employ an elicitation process which is easy to understand and consumes as little time as possible.

A less time-consuming alternative to manual elicitation is utilizing a parameterized model, where the NPTs are constructed according to a formula whose variables are elicited from experts. One common method is using NoisyOR operators [67][68], which provides a logarithmic reduction in the amount of variables required for describing the state distribution of a child node. A NoisyOR operator naturally ignores any interaction between the parent nodes, however, it was determined during the first stage interview that no significant interaction exists between the risk nodes. Using a NoisyOR operator for describing the state of a child node is however only practical if all nodes are Boolean. For the purpose of multivalued nodes, a generalization named Noisy-MAX exists. However, estimating the variables of a Noisy-MAX operator becomes increasingly difficult and inaccurate with a large amount of parent nodes.

Another method used by Fenton et al. [55] includes estimating the distribution of a node's value using the truncated normal distribution. In this method, the average and variance of the distribution are defined as a function of the node's parent's values, where the function is defined by the experts. This general method provides a means to elicit large NPTs quickly and relatively effortlessly but can only be used with so called ranked nodes which represent an abstraction of a continuous variable. Even if the underlying variable of the aforementioned 3-step impact scale could be represented as a continuous variable, the inclusion of the fourth step *negligible* does not allow this representation.

For the purpose of this risk assessment, a method was designed with the objective of being easy to understand even by experts with little or no statistical experience. The resulting method consisted of assessing each risk-consequence-pair individually and thereafter combining this information into a NPTs using a tool discussed in section 4.1. The process is visualized in figure 10 and described in more detail below. Due to practical and technical reasons, one of the interviews only followed this process approximately.

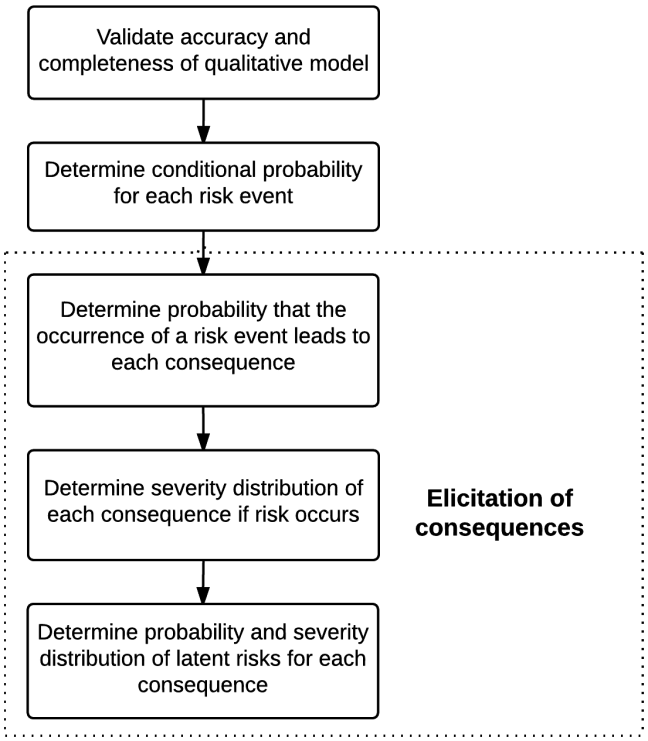


Figure 10: Expert elicitation stage 2 interview process

First the interviewees were shown qualitative model built based on data gathered during the first stage interviews. Figure 11 in section 4 shows a subset of the model, which describes the risk events that can cause leakage of personal data. The experts were also provided with the documents, included in appendix D, defining and illustrating the different risk events, consequences and their severities. The experts

were then asked to validate the completeness and accuracy of the model including the existence of relationships between the nodes. During this validation stage, the number of nodes did not increase or reduce but one additional causal relationship between two risks was added. Second, the experts were queried about conditional probabilities of risks with direct mutual relationships. If statistical data was available for the marginal occurrence of the risk, this value was used as a baseline for the query. The result for each node was a *1x2* or *2x2* NPT of probability values, depending on whether the node had a parent or not.

Next, each risk-consequence-pair was assessed individually and for each pair, the following two questions were asked: (1) how likely is it that occurrence of the risk leads to this consequence, (2) if the risk occurs and leads to this consequence, how severe are the most likely effects, i.e. how strong is the impact. To elaborate on the second question, the experts were also asked how likely it is that the impact falls into each of the impact strength categories (low-medium-high). The interviewee's were provided with the example distributions of possible impact strengths shown in appendix A to ease the process. Typical responses from the experts included choosing one example distribution and describing a small change to the distribution with which it would represent the expert's opinion.

For each consequence, the experts were also asked whether the risk event nodes in the model contribute 100 % of the respective consequence or whether other relevant risk events, i.e. latent risks, also exist. If all relevant risks were not portrayed by nodes in the model, the experts were queried for the combined probability of occurrence and consequence severity distribution of the latent risks.

This method was easy for experts to understand and follow, which minimized the time and effort required for familiarizing the experts with the process. However, the high amount of risk-consequence-pairs caused the method to be more time consuming than parameterized methods such as NoisyOR [67] or the method which utilizes a truncated normal distribution [55]. As a counterbalance to the higher amount of time required, the method used here is not as prone to the typical loss of accuracy when representing large amounts of variables with a simplified function. Compared to manual elicitation of NPTs, this method still provided a nearly hundredfold reduction in variables to be elicited.

3.4 Results from expert elicitation

3.4.1 Risk events

The original risk list used in the stage 1 interviews included a separation between three different types of malware, which were (1) spyware, (2) diallerware and (3) other types of malware. However, interviewee's found it easier and more logical during the first stage to consider all malware as one risk category. The main difference between these malware types concerns the consequences of an infection, wherein a separation is better made at consequence level as will be defined in subsection 3.4.2. Thus the three types of malware were combined into one risk named simply *malware*.

The definition of the risk *shoulder surfing* was broadened to also cover eavesdropping on voice calls in the physical proximity of a user, i.e. not on the network. The risks originally named *client side code injection* and *premium number scam* were renamed to *remote code injection* and *premium number fraud* in order to better represent the risk definitions.

The risk *surveillance* was originally defined as targeted surveillance on a device, for example by installing hidden spying tools on it. However, the updated risk list includes malware, which covers these tools, and on the other hand surveillance on network level. Thus, the original *surveillance* was deemed unnecessary. Technical failures of the network and device, including software crashing, arose as a risk in several first stage interviews and the risks *technical failure of device* and *technical failure of network* were thus added to the list.

Based on the above and other information received from the experts during the first stage interviews, the risk list presented in subsection 3.2.2 was updated. Further, in order to identify the most important risks from this list, each risk event was given points based on how important the experts viewed it as on average. Each expert was asked to choose a maximum of 10 most important risk events and order these based on total importance taking into account both probability and impact. Each risk was then given points (1-10) based on how high the risk event was on each expert's list. The risks were ordered based on total points, hereafter referred to as importance score. This list of risks including each risk's importance score can be seen in table 6. The risks are described in more detail at the end of this subsection. Other risks not in the scope of this thesis are briefly discussed in section 3.4.3.

#	**Risk event name**	**Importance score**
1.	Loss or theft of device	86
2.	Unintentional data disclosure	85
3.	Malware	79
4.	Technical failure of device	58
5.	Unauthorized physical device access	50
6.	Network device spoofing attack	42
7.	Vendor backdoor	30
8.	Surveillance on cellular network level	30
9.	Premium number fraud	27
10.	Technical failure of network	23
11.	Shoulder surfing or eavesdropping	20
12.	Phishing	13
13.	Sniffing on legitimate networks	11
14.	Mobile payment systems abuse	8
15.	Remote code injection	8
16.	Ad tracking system abuse	6
17.	Attacks on decommissioned devices	4
18.	Compromising wireless encryption	4
19.	Denial of Service	3
20.	Remote wipe	2
21.	Data mining from cloud uploads	2
22.	Cloning SIM card	2
23.	Resource abuse	2

Table 6: Final list of risk events and their respective importance scores

The risk importance scores shown in table 6 exhibit a long tail, wherein the four most important risks represent approximately 50 % of all points. In order to build a Bayesian network of most important information security risks and consequences related to smartphone use, the list in table 6 was narrowed down based on importance. The interviewed experts were all able to name their subjective most important 1.-4. risks quite easily, while 5.-10. were more difficult to identify and order. Therefore, the risks for further evaluation were chosen so that all such risks were included, which were identified by an expert in their respective top 4 risks. The resulting 13 risks are presented in the upper part of table 6 and detailed descriptions of the risk events can be found below.

Loss or theft of device

Loss or theft of device includes all scenarios where a user's smartphone is lost or stolen but excludes scenarios where the device is temporarily used by an unauthorized user and then returned, which are covered by the risk *unauthorized physical device access*.

Unintentional data disclosure

Unintentional data disclosure includes all scenarios where a user unknowingly or unintentionally discloses data through a legitimate application. In this context, legitimate is defined as something the user has agreed to in the terms and conditions of the application.

Malware

This threat includes all malicious applications that can run on smartphone such as spyware, adware and diallerware. This threat does not include applications which perform legitimate actions such as ad tracking without the user's knowledge, if the user has agreed to these actions in the application's terms and conditions.

Technical failure of device

This threat includes any technical failure of device hardware or software which significantly affects the device user in any way. Significant effects can include for example data loss or unavailability for a significant amount of time. Due to the complex nature of smartphones, a significant technical failure of the device is difficult to define unambiguously. The interviewed experts were thus allowed to use their own judgement when evaluating this risk.

Unauthorized physical device access

Unauthorized physical device access includes all scenarios where an unauthorized person accesses the smartphone locally. A prerequisite for this risk being realized is that the unauthorized person is able to bypass the potential logical access control on the device. However, this risk excludes all unauthorized access by family members. The reasons behind the exclusion were the relatively high occurrence and low impact of these situations compared to scenarios where a less familiar person accesses the device.

Network device spoofing attack

Network device spoofing attack includes all events where a user's smartphone connects to a rogue device that impersonates a trusted device such as a WLAN access point or cellular base station. Depending on the technology, the connection can be formed automatically or require action from the user. By definition, the rogue network device must have been deployed with malicious intent, e.g. to monitor or intercept data from users.

Vendor backdoor

Vendor backdoor as a risk event describes events where a backdoor or information collection agent collects data from or controls a smartphone. By definition, the backdoor or information collection agent was inserted into the device by the device or OS vendor, either before sale of the device or afterwards as an update. Other types of malicious software are covered by the risk *malware*. The vendor backdoor risk is mainly speculative, since evidence does not exist to confirm the existence of widely spread backdoors in smartphones.

Surveillance on cellular network level

Surveillance on cellular network level is defined as surveillance or tracking performed on the network level, independent of a user's device's security. Methods for performing such an attack include accessing a network base station, an operator's core network or the worldwide SS7 signalling network. For the SS7 signalling network, any insecure network operator could act as a malicious entry point and allow the attacker to surveil users in other networks.

In addition, network operators are able and likely to collect a significant amount of data about their users. Some of this data can also be required by law enforcement to be stored for a certain period of time. While this data is stored with the network operator, it could be accessed and misused by a dishonest employee, an outside intruder or a foreign government, for instance.

Premium number fraud

Premium number fraud includes events where billing is caused by calls or messages to premium rate services without the user's approval and consent. This risk includes both scenarios where a user is tricked into calling or sending a message to a premium rate service as well as scenarios where malware performs the action without the user's knowledge.

Technical failure of network

Technical failure of network includes all cellular network failures that significantly affect smartphone use for at least 1 hour. A network failure can for example cause unavailability of the whole cellular network, unavailability of data connections or significantly decreased connection quality. The risk event is defined to last at least 1 hour in order to exclude small network failures that cause negligible inconvenience to users but happen relatively often. A Bayesian networks risk assessment of Finnish mobile network availability found that the most significant triggers leading to unavailability were natural disasters, cyber-attacks and bad operation by own employees [76].

Shoulder surfing or eavesdropping

Shoulder surfing or eavesdropping includes purposefully watching, recording or eavesdropping on a user while they use their smartphone. However, this risk excludes surveillance performed by family members due to its relatively small impact but high occurrence.

Phishing
Phishing includes all scenarios where a user gives personal or confidential information to an attacker through the smartphone unknowingly.

Sniffing on legitimate networks
Sniffing on legitimate networks includes scenarios where a user's data traffic is monitored or intercepted by a third party on a legitimate network. Scenarios where the attacker has control over the whole network are described in threat *network device spoofing attack*.

Mobile payment systems abuse
Mobile payment systems abuse includes attacks that target a payment system used on smartphones such as Elisa Lompakko or Danske Mobile Pay, or a mobile banking service.

Remote code injection
Remote code injection includes all scenarios where a smartphone interprets an untrusted input as commands. Untrusted inputs include instant messages and websites, for example.

Ad tracking system abuse
Ad tracking system abuse includes all attacks which target an existing ad tracking system. Aggressive ad tracking libraries collect extensive amounts of personal information and sometimes leak this information even in plaintext [71]. Information collected by ad tracking libraries such as Burstly include sexual orientation, political affiliation, number of children and income [71].

Attacks on decommissioned devices
Attacks on decommissioned devices include scenarios where an outsider recovers data from a device that has been decommissioned or recycled without properly wiping its memory.

Compromising wireless encryption
Compromising wireless encryption includes scenarios where an attacker is able to read or modify the wireless encrypted data traffic between a smartphone and a cellular base station.

Denial of Service
Denial of Service includes all network-based DoS attacks that target smartphones. Unavailability caused by malware on the device is covered by the risk *malware*.

Remote wipe
Remote wipe includes scenarios where a smartphone's memory is emptied using a remote wipe functionality without the user's knowledge or approval. Many modern smartphones include remote wipe functionalities automatically available for example through Android Device Manager. Also, employer-provided smartphones may have remote wipe systems administered by the employer.

Data mining from cloud uploads
Smartphone users often upload large amounts of data to cloud services. Some uploads are deliberate such as image uploads to social networks while other information might be uploaded by an application on the device without the user's knowledge. Even if any one piece of information would not pose a security risk or comprise a security breach, modern data mining methods can process this data in order to gain information whose leak would fulfil the definition of a security breach.

Cloning SIM card
Cloning a SIM card in order to impersonate another subscriber on the cellular network.

Resource abuse
Resource abuse includes all threats which intentionally abuse a smartphone's resources such as the battery or CPU.

3.4.2 Consequences

One substantial change was made concerning the initial consequence categories described in subsection 3.3. The sixth consequence category, *access to other devices or services*, was identified to be affected by the same threats as the *leakage of personal data*. Thus the separate consequence category was removed and its contents were merged into the category *leakage of personal data*.

In addition to the original consequence categories, experts also brought up consequences that affect a user's reputation. In this definition of consequence categories, reputation damage caused by the leakage of personal information is covered by the *leakage of personal data*. On the other hand, reputation damage could also be caused by impersonation of the user through a device or service used on the smartphone, for example by sending an SMS using the user's subscription. These consequences were originally included in the category *access to other devices and services* and thus now a part of *leakage of personal data*. Therefore, a separate consequence category was not deemed necessary for reputational risks.

The final consequence categories are described below.

Leakage of personal data
Leakage of personal data includes all scenarios where non-public information related to a user's personal life, as opposed to their work, becomes available to an attacker or a third party. Also, this category includes scenarios where an attacker or a third party gains access to a service used by the legitimate user.

Leakage of confidential data
Leakage of confidential data refers to scenarios where information related to the user's employer becomes available to an attacker or a third party.

Data loss or corruption
Data loss or corruption includes all situations where the user permanently loses access to data or cannot anymore utilize the data due to corruption. This category also includes scenarios where the data would be technically possible to recover, but is not recovered due to the user's lack of knowledge or resources.

Financial consequences
Financial consequences include all direct effects on the user's or device owner's finances. This category does not include financial consequences experienced by the user's employer due to confidential data leakage or indirect consequences to the user such as loss of income due to employment termination caused by the careless actions that led to the data leakage.

Unavailability of device or services
Unavailability of device or services includes all situations where the user does not experience full availability and functionality from the device and its services as compared to a typical baseline availability and functionality experienced using the same device and network.

3.4.3 Other notions

The interviewed experts also raised other concerns related to smartphone use, which were not in the scope of this thesis and thus not described in subsection 3.4.1.

One such risk was that of driving while using a smartphone, e.g. talking or texting. Although this risk can have very severe consequence, it was not classified as an information security risk since neither the risk event or its consequences relate closely to information. Another physical safety risk was the possibility of a violent theft. Expensive smartphones are popular among thieves and using an expensive smartphone in public could thus theoretically lead to the user becoming the victim of a violent robbery, which could danger the user's health.

A technical risk brought up in the interviews goes by the name of *BadUSB* [72]. This threat exploits a fundamental weakness in the Universal Serial Bus (USB) standard which allows USB devices to maliciously monitor or control a computer. Android phones are one of the devices most suitable for use as an attack vector. This threat was determined to be out of scope of this thesis, because traditional computers are the attack target and smartphones are merely a tool for executing the attack.

4 Bayesian network modelling

4.1 Modelling methods

The qualitative model, i.e. the graphical network of risks, consequences and their dependencies, was built based on the consolidated results from the first stage interviews. A subset of the complete model can be seen in figure 11. The choice of risks to be included in the network was made based on importance according to the method described in subsection 3.4.1. All dependencies which were described as significant by at least one expert, were included in the initial model. The dependencies that had been described as theoretical and insignificant were left out. When the insignificant dependencies had been removed, the model did not include any loops to eliminate.

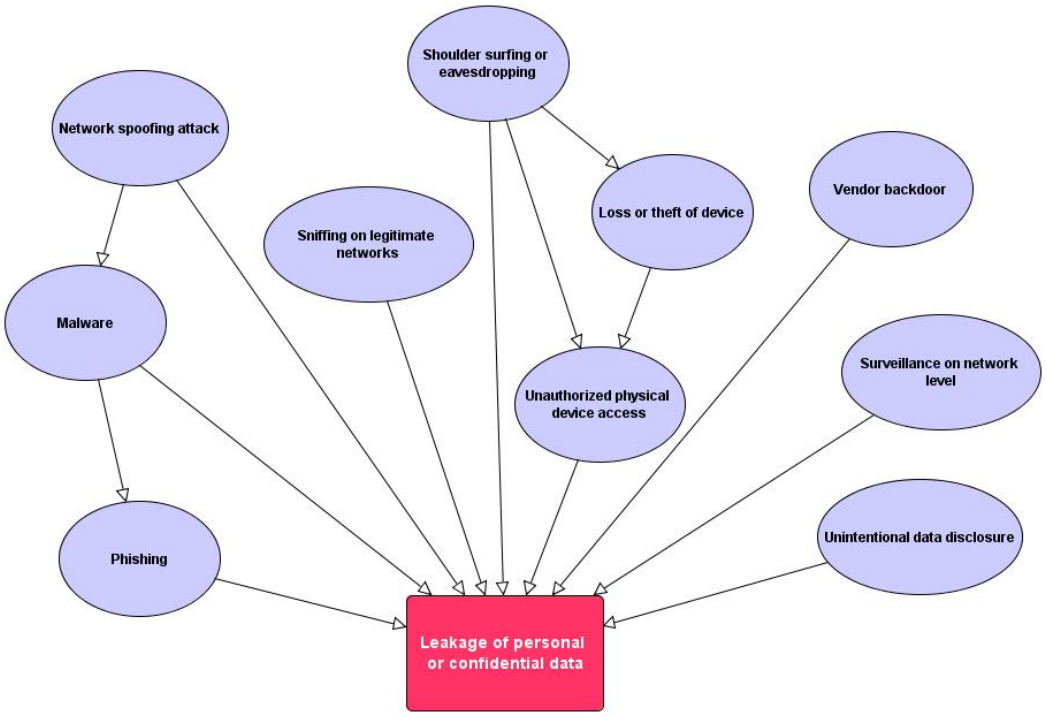


Figure 11: Qualitative model of data leakage consequence and its causes

For the quantitative part of the model, i.e. the NPTs, an arithmetic average was taken of the probability values and severity distributions given by different experts. A Microsoft Excel based tool, described in appendix C, was designed to generate NPTs based on these averages. The possibility of using a weighted average in order to emphasize specific experts' responses was also considered. However, due to the experts' different areas of expertise, determining a weight for each expert's answers was deemed unjustified.

A parameterized model was also considered as an alternative to the tool-based NPT generation method. However, as described in section 3.3.3, a suitable parameterization method was not found.

4.2 Bayesian network model

The Bayesian network constructed in this thesis describes the most important risk events and consequences related to smartphone use in Finland during year 2015. Probabilities in the network describe the probability of an event or consequence for a single smartphone user during one year. All probabilities, dependencies and effect strengths have been determined without making any extra presumptions such as device model, OS or user age.

Figure 12 illustrates the risk events and consequences related to *shoulder surfing or eavesdropping*, with other risks hidden for clarity. In addition to the possibility of directly causing leakage of confidential or personal data, *shoulder surfing or eavesdropping* can also lead to theft of the device, for example a scenario where a thief steals the device after seeing its passcode. Furthermore, the thief might be interested in accessing the device's information and services, thus realizing the risk *unauthorized physical device access*, as opposed to wiping the device's memory and selling it.

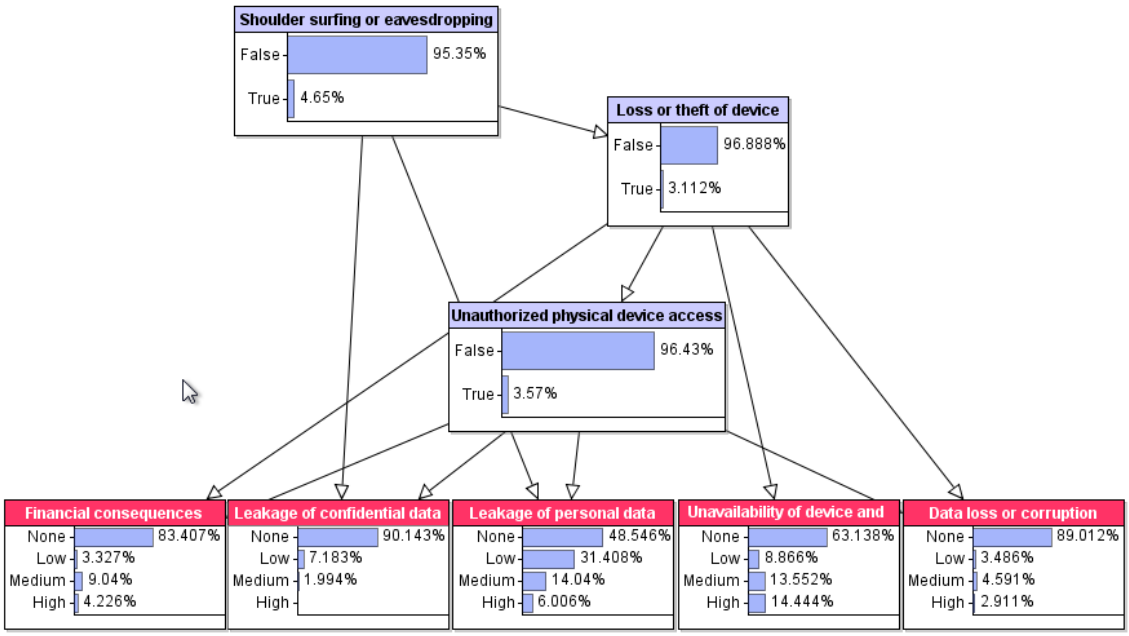


Figure 12: Bayesian network model demonstrating the risk *shoulder surfing or eavesdropping*. The figure depicts the combined consequences of all risk events in the complete model, but however, most risks are hidden in this figure for clarity.

Figure 13 shows the complete Bayesian Network structure including the nodes' state probability distributions. As can be seen from the figure, most risk events in the network relate to at least two consequences and some also to other risks. However, the majority of the risks are not directly dependent on other risks included in the network.

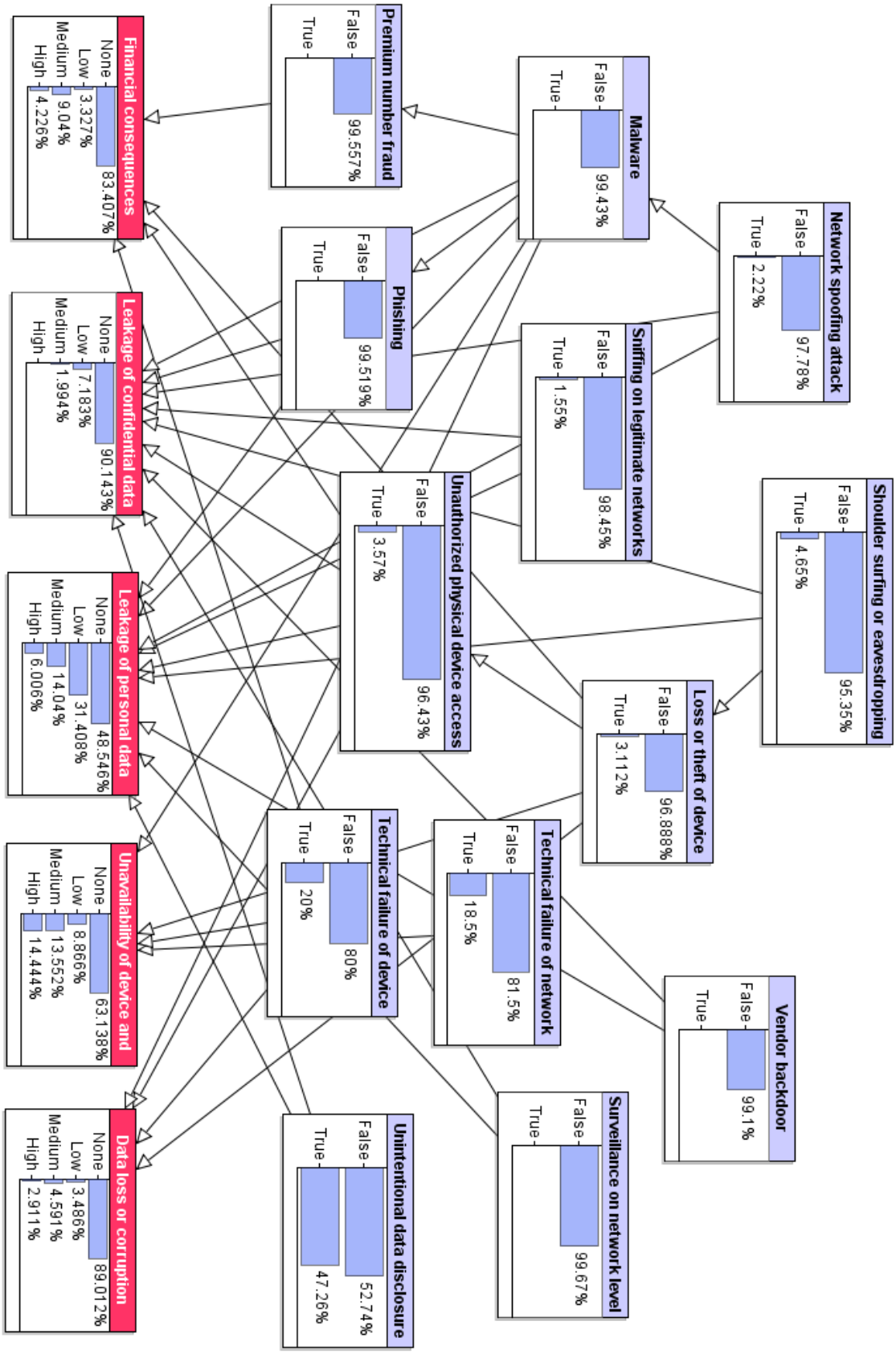


Figure 13: Complete Bayesian network model of information security risks related to smartphone use

4.3 Analysis using Bayesian network model

Figure 14 visualizes the probability of occurrence for each risk, which shows that most risks are unlikely to occur during one year. However, based on figure 14, experts believe that almost 50 % of smartphone users become victim to *unintentional data disclosure* during a year, i.e. share more information to other parties through their smartphones than they acknowledge or would be willing to share.

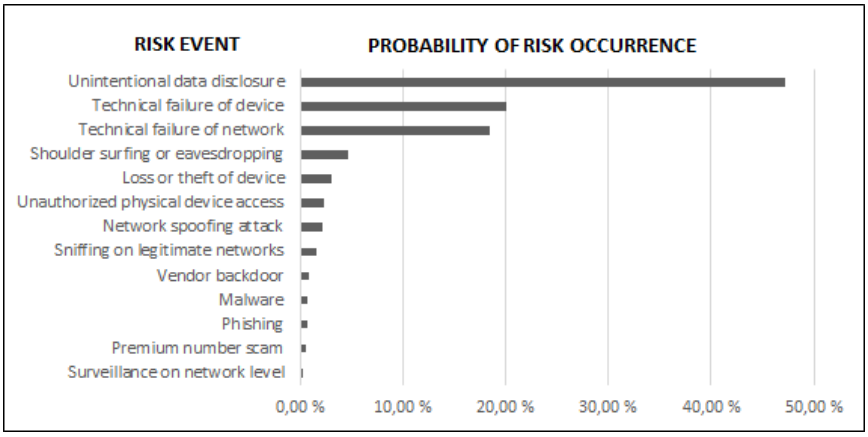


Figure 14: Probabilities of occurrence for each risk

Figure 15 visualizes the probability and severity distribution of consequences when a risk event occurs. Based on the figure, the risks *unintentional data disclosure* and *vendor backdoor* are both very likely to cause leakage of personal data when they occur but the effects of *vendor backdoor* are more likely to be severe.

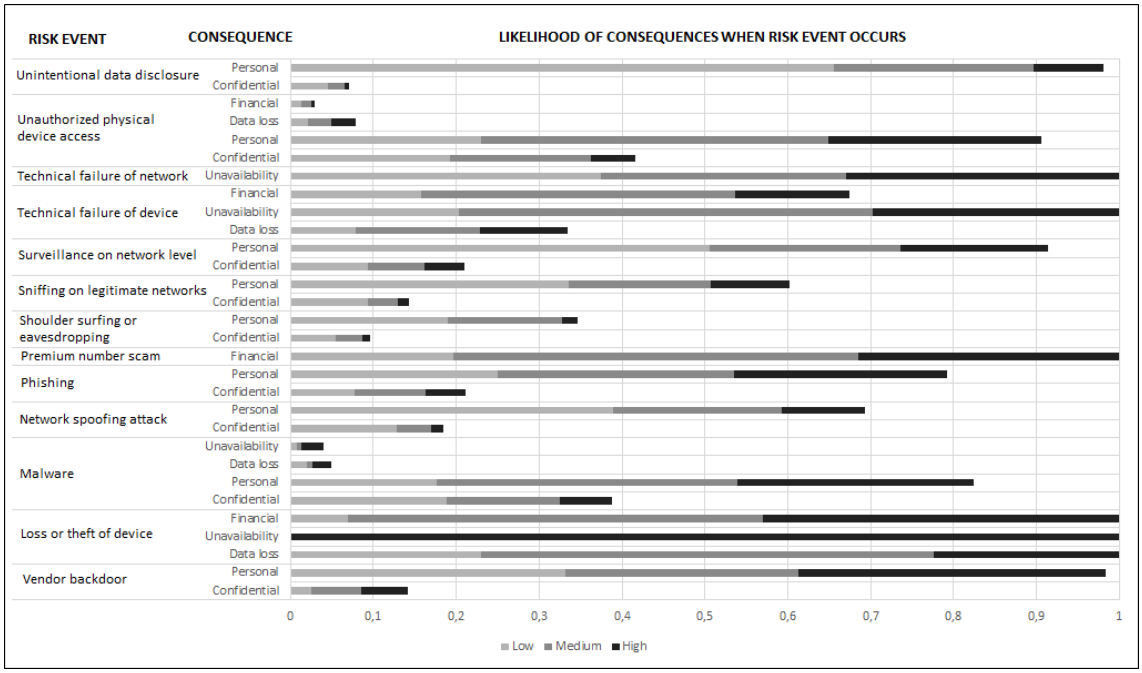


Figure 15: Consequence distribution of each risk event

Figure 16 describes the probability distributions of data leakage consequences of different severities. From the risk assessment results, it seems clear that a data breach is considerably more likely to concern a smartphone user's personal information than confidential information related to the user's employer. One likely reason for this is that employer-provided services most often use encryption and strong authentication. Although businesses and governmental entities usually have an incentive to ensure the security of services used, personal end-users' choices rely more on other factors such as price and ease of use. On the other hand, employer-provided smartphones can also have built-in security restrictions which limit exposure to any type of vulnerabilities.

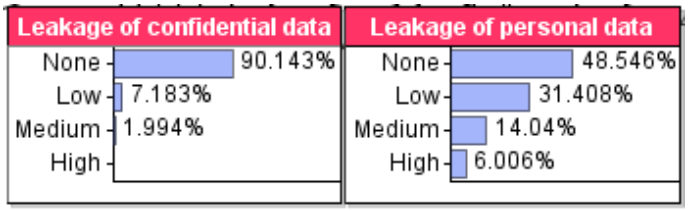


Figure 16: Data leakage severity distributions.

Figure 17 shows a sensitivity graph describing the effects of individual risk events on medium- or high-severity confidential data leakage. According to the analysis, the consequence is most sensitive to occurrence of *unauthorized physical device access* or *malware*. This is reasonable as an unauthorized user would have access to all services which do not require additional authentication and malware with elevated access could access all data on the device and monitor interaction between the device and user. However, most devices used for confidential purposes should require a passcode for unlocking the device, which might not be sufficiently represented in the results.

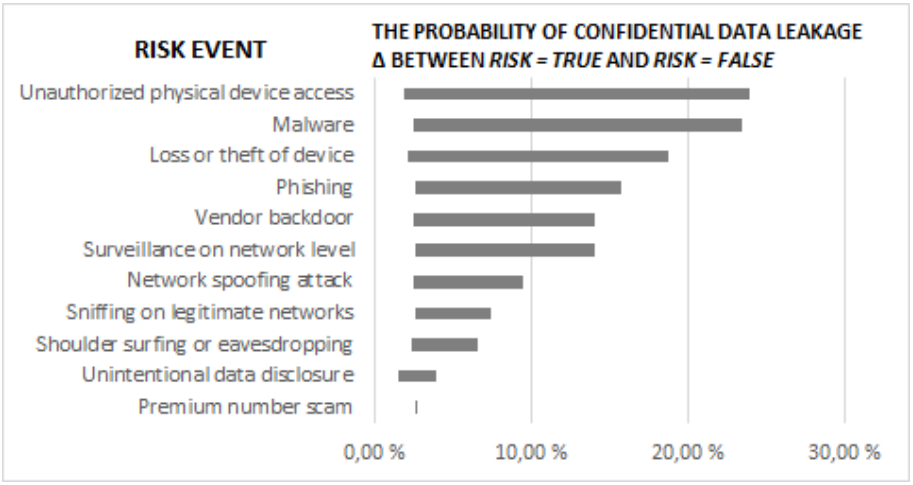


Figure 17: Effect of occurrence of individual risk events on medium- or high-severity leakage of confidential data

Figure 18 represents the effects of individual risks on medium- or high-severity personal data leakage. The results resemble those of confidential data leakage in figure 16. However, two clear differences exist between these results: (1) all risk events have a significantly higher probability of affecting personal data than confidential data and (2) the effect of *unintentional data disclosure* is much higher relative to other risks' effects on leakage of personal than confidential data.

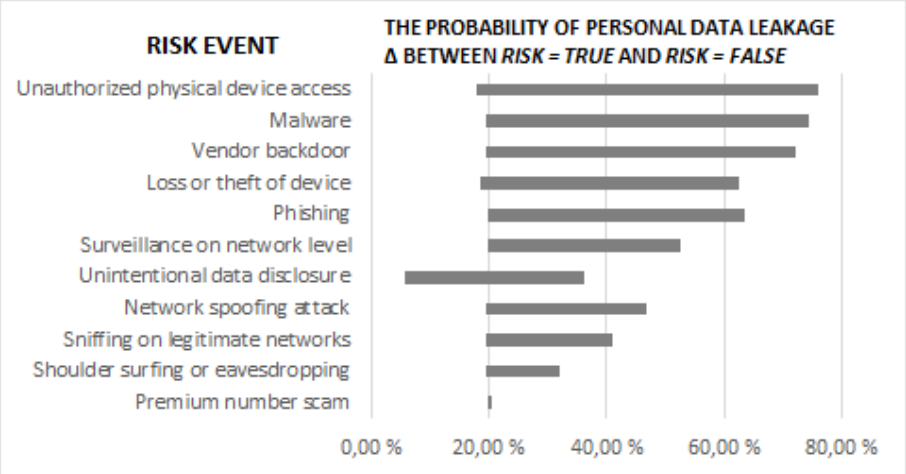


Figure 18: Effect of occurrence of individual risk events on medium- or high-severity leakage of personal data

Figure 19 shows the severity distribution of personal data leakage in a scenario where it is known that *unintentional data disclosure* has not happened. According to the created model, the majority of low-severity occurrences are caused by unintentional disclosure of data by the user. However, a considerable probability of personal data leakage still remains, especially for medium- and high-severity events.

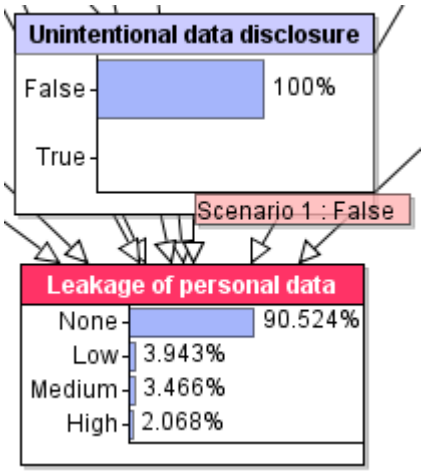


Figure 19: Leakage of personal data when *unintentional data disclosure* is set to *false*

Loss or corruption of data related to smartphone usage seems to be a relatively unlikely scenario as can be seen in figure 13. Figure 20 shows that the risks leading to data loss or corruption are most often *loss or theft of device* or *technical failure of device*.

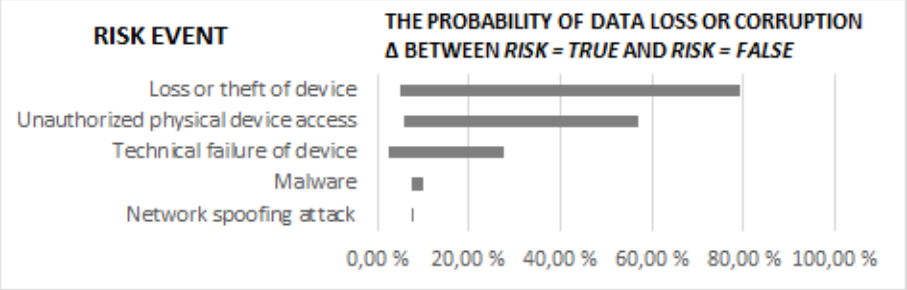


Figure 20: Effect of occurrence of individual risk events on medium- or high-severity data loss or corruption

According to the model, low-severity and medium-severity unavailability is very rarely caused by anything else than technical failures of the device or network. On the other hand, high unavailability is rarely caused by anything else than *loss or theft the device*. The effects of individual risks on medium- or high-severity unavailability can be seen in figure 21.

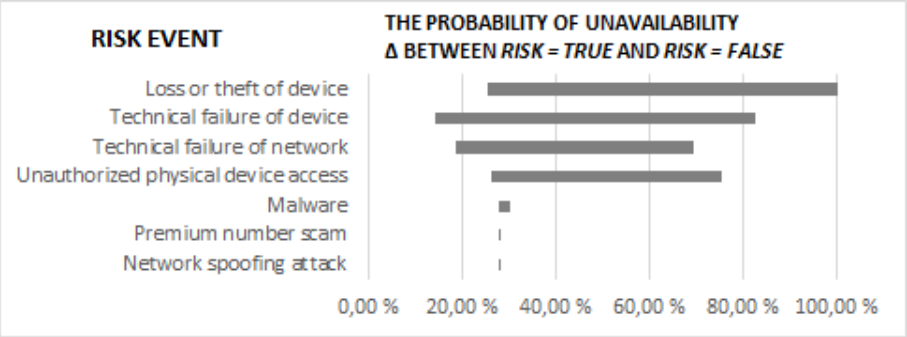


Figure 21: Effect of occurrence of individual risk events on medium- or high-severity unavailability

The sensitivity graphs in figure 17, 18, 20 and 21 show that a significant amount of smartphone information security incidents are caused by risks which are dependent on insecure action by the user, such as *unauthorized physical device access*, *malware* and *phishing*. Users clearly have an important role in securing their smartphone use and therefore, there is a need for more security awareness among smartphone users.

4.4 Controls and mitigants

The Bayesian network visualized in figure 13 in subsection 4.2 defines from an information security perspective the risk events and consequences related to smartphone use. In addition, the network can be extended with controls, which affect the probability of the risk event, and with mitigants, which affect the probability and severity of consequences when the risk event occurs. Figure 22 shows an example where careful use of WLANs decreases the probability of becoming victim to *sniffing on legitimate networks* and careful use of mobile applications decreases the severity of consequences in such a scenario.

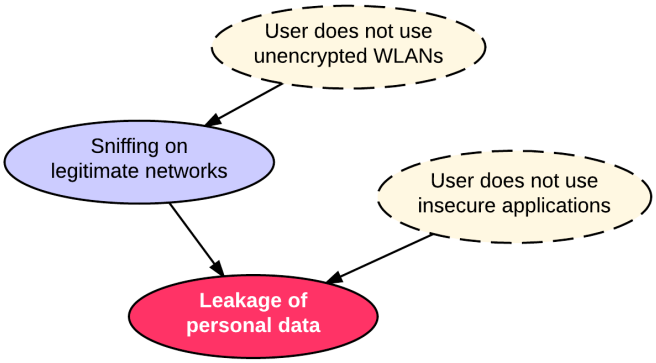


Figure 22: Risk event and consequence with example control and mitigant

Some of the most important controls and mitigants identified during the risk assessment process are discussed below.

Use of insecure applications
Use of insecure applications significantly impacts the consequences of becoming victim to a *network device spoofing attack* or *sniffing on legitimate networks*. If the operating system and all applications on the device only use secure APIs and strong encryption for communication, an attacker should not easily discover sensitive information. Also, if reliable two-way authentication is used, a Man-in-the-Middle attack should also not be feasible. In addition to these two risks, badly designed application security can also benefit malware which could potentially gain access to another application's data or functionalities. Use of insecure applications, or lack thereof, can thus be viewed as a mitigant to the aforementioned risks.

Device updates
Another similar mitigant is how quickly updates are installed on the device when vulnerabilities are discovered. The speed of update process is often dependent on the operating system developer, the device vendor and the user as well as possible other quarters.

Use of unencrypted WLANs
The risks *sniffing on legitimate networks* and *device spoofing attack* are most

feasible if the user connects to unencrypted WLANs. The use of unencrypted wireless networks can thus be seen as a control to these risks.

Use of cloud services
Cloud services for smartphones often offer extensive backup functionalities and can thus mitigate data loss consequences. However, cloud services also introduce new threats against the confidentiality of this information. In addition to the possible vulnerabilities in the data storage solutions and the user account security, a cloud service provider could also willingly share the information with third parties such as governmental agencies.

Device OS
Different smartphone operating systems vary with regards to their security measures and application architecture such as described in subsection 2.1.1. According to research by Verizon [64], most of today's malware is targeted at Android users and the use of an Android device can thus be viewed as a control to the *malware* risk event. This is likely mostly due to the relatively open policies of Google Android's Play Store described in section 2.1.1. In contrast, most insecurities in legitimate applications are found on iOS applications [71], which can be viewed as a control to risks such as *phishing* and *sniffing on legitimate networks*.

Installation of applications outside the official application store
Although some malware exists on all official application stores, most of the malware in circulation exists outside these [75]. Prohibiting installation from other sources can thus work as a control to the risk *malware*. On Android devices, this is a setting which the user can toggle on and off. On standard iOS devices, this is not possible and requires "jail breaking" the device.

Based on the controls and mitigants discussed above, the following high-level recommendations could be given:

- A user should ensure that their smartphone is always updated to the newest software version. Also if possible, users should choose devices whose vendors release security updates quickly when vulnerabilities are discovered.
- The use of unencrypted WLAN's should be avoided. If one must be used, a secure VPN connection is recommended.
- Cloud services are an easy solution to backing up data from smartphones, but a user should consider its privacy and confidentiality implications before use.
- Smartphone users should install applications only from the official application stores and disable installation from other sources.

In addition to the aforementioned variables, many other controls and mitigants exist which depend on the user's actions. In essence, the user's knowledge and understanding of choices and their respective security implications have a significant impact on the security of their smartphone use.

5 Discussions

5.1 Assessment of results

The Bayesian network model built during this risk assessment and described in subsection 4.2 seems realistic as a representation of the actual risk space surrounding smartphone use. The results do not include any significant surprises but give more insight into the importance of different risks and their consequences. The author has not found any similar prior research.

The information elicited from experts is mostly well in line with that found in statistics and existing research. For example, the conditional probability that the data on a lost or stolen smartphone is accessed, reflects the results of Symantec's Honey Stick Project [83]. Some risks also exhibit a high variance between different expert's answers, such as the *unintentional data disclosure* risk, where estimates of probability ranged from less than 1 % to 95 %. However, the average of answers received from the 8 experts seem realistic. One special threat that arose in this risk assessment is the speculative *vendor backdoor*, for which the experts had very different opinions. According to some experts, at least 10 % of the smartphones in Finland are likely to include an active backdoor designed by the device vendor or OS developer, whereas some experts believe that none of the smartphones used in Finland have such a backdoor.

Some experts were also uncomfortable estimating probabilities without any statistical data. Data leaks regarding company confidential information is one aspect where experts felt particularly insecure due to lack of information. It was also brought up that businesses do not necessarily report all data breaches to outside parties due to possible reputational damage.

The constructed model is essentially a document of the 8 interviewed experts' average opinions, which were elicited using a process designed to introduce as little bias as possible. Due to the high variance in the experts' opinions, it is reasonable to assume that a similar process with a larger amount of experts could have generated slightly different results. A biased sample of experts, such as choosing all experts from one organization, would most likely have caused significant bias in the results.

5.2 Exploitation of results

The Bayesian network model can be used as is to illustrate the significance of different smartphone risks in different scenarios. For example, the most important risks are very different when a user is concerned with leakage of data than with loss or corruption of data. The model can be of use to security professionals who need to identify and communicate relevant risks related to smartphones.

The model could be applied to other countries by updating the parameter values and possibly also the structure to represent the local environment. Some risks such as the *mobile payment systems abuse* did not reach the top 13 risks in this assessment due to their scarcity of users in Finland. However, in a country where these services are more popular, the risk would likely merit an individual node in the model.

The use cases of smartphones are still evolving rapidly and thus the risks of smartphone use change constantly. In order to ensure validity of the model, the parameter values as well as the structure should be updated regularly. The parameter values should be updated at least yearly and the structure should be re-evaluated each time the use cases or technical features of smartphones undergo significant changes.

5.3 Future prospects

A common expectation between the experts is that the mobile risk space is going to undergo significant changes in the near future. One opinion is that the ever-strengthening connection between smartphones and the physical world will cause users to carry around more and more sensitive information on their phones. With the growing number of mobile banking and mobile payment systems, the amount of potential targets for financial attacks on smartphones is also quickly rising. Although attacks on mobile payment systems or banking services were not identified as a large threat at this moment in Finland, the interviewed experts expect these incidents to become more common in the future.

As discussed in section 2.1.1, smartphones often include a large amount of sensors that collect different data. Malicious access to this data is becoming increasingly threatening as more information is continually collected through these sensors. Risks such as *malware* and *vendor backdoor* present a threat that could continuously transfer the sensor data to another party.

According to a 2015 data breach report [64], an insignificantly small part of reported data breaches involve mobile malware or smartphones. On the other hand, traditional malware is prevalent and still evolving. There is no clear reason why the focus of attackers would not shift towards mobile such as the focus of users has done.

While mobile phishing is not yet a large problem, research shows that many services used on smartphones are vulnerable to sophisticated phishing attacks [73]. The interaction between web sites and mobile applications as well as interaction between applications could in most cases be spoofed by attackers in an unsuspicious way, leading to undetected attacks.

5.4 Further development of model

In the current model, the consequences such as *leakage of personal data* do not specify the concrete effects they have on a user's life. As further development of the

model, the risk events and consequences could be divided into more specific events and effects such as *reputational damage* and *indirect financial consequences*. Also, the risk events in the final model are not necessarily individual risk events but rather categories of risk events, such as *loss or theft of device*. The model would be more informative but also more complex, if it made a distinction between each concrete event such as loss and theft of device.

Another possible extension of the network would be to add demographic parameters such as age, occupation or income of the user. With demographic parameters, the network would give more insight into an individual user's vulnerabilities as well as what kind of security awareness is lacking with each demographic group. Extending the network with demographic parameters could allow the model to be used in insurance to determine the probability and thus appropriate insurance premium for scenarios which cause financial consequences such as loss or theft of the device. Smartphone vendors and operators could use the model to deliver highly targeted security awareness information to users.

The model could also be extended with controls and mitigants, such as those described in subsection 4.4. Such a model would give insight into what causes the occurrences of risk events and how these can be prevented. The extended model could be used for identifying the most dangerous practices of smartphone usage as well as which security measures would be most useful for prevention or mitigation of risks. By determining the costs of different security measures, a cost-benefit analysis of controls and mitigants could also be performed. Such a model would be beneficial to most organizations' IT departments, for instance. Other potential value networks could describe the viewpoint of an attacker, who wishes to gain money with minimal investment and risk, or a consumer, who wishes to avoid security incidents with minimal cost. However, quantifying potential damages is difficult due to the lack of available data.

5.5 Evaluation of risk assessment process

Based on this risk assessment, Bayesian networks seem suitable for information security risk assessment. The method produces a very flexible model which can be used for various kinds of analysis such as sensitivity analysis. However, the process for gathering the necessary information and constructing the network is time-consuming and poses challenges for networks with a large amount of nodes. There is clearly room for improvement in the Bayesian network tools provided by vendors for elicitation and consolidation of expert opinions. Efficient tools which ease this process could make Bayesian networks considerably more practical.

The difficulty of the information gathering process also depends on how much prior knowledge exists concerning the domain. Existing statistical information concerning the probabilities of different events is especially helpful for construction of Bayesian networks.

One of the aspects that Bayesian networks are especially suitable for, is taking into account and analysing the effects of interaction between several risk nodes. However, in this case most risks only had a direct causal relationship the consequences. Thus this assessment did not unveil the full potential of using Bayesian networks for risk assessment.

The expert elicitation method designed during this study is seen to be suitable for the purpose. The method is easy for experts to understand and follow, which is of essence when interviewing busy domain experts. However, the large amount of risk-consequence pairs caused some visible fatigue of the experts during interviews. Based on this experience, elicitation sessions longer than the two-hour-sessions held during this process cannot be recommended. On the contrary, it could be beneficial to split the sessions into even shorter intervals with at least short breaks in between.

The total time required with each expert, 4 hours, is reasonable for the purpose of this risk assessment. However, the time required with experts increases linearly with the amount of nodes in the network. Therefore, this method could easily become too time-consuming for elicitation of larger networks. Although this method is more time-consuming than a parameterized method, it provides more accurate results. There is still clearly potential for both researchers and Bayesian network tool vendors to develop less time-consuming methods of consolidating expert opinion.

The process generates results that are easy to interpret and utilize to build the Bayesian network model. However, while the generation of the Bayesian network model from interview results is mostly automatized, any interaction between risk nodes must be manually taken into account.

6 Conclusions

The purpose of this thesis was to perform an information security risk assessment of smartphones using Bayesian networks. Most information was gathered during a two-stage expert elicitation process, in which 8 domain experts were queried for the relevant information security risks, their causal relationships, consequences and quantitative probabilities. The experts represent various experience, knowledge and viewpoints related to information security, thus ensuring the completeness of the elicited information. The expert interviews followed a process designed as part of this thesis in order to facilitate accurate and simple elicitation.

The outcome of this thesis is a Bayesian network model which documents the information security risks related to smartphone use and can be extended with new data when available. The model shows that the most important risks in Finland include traditional information security risks such as *malware* and *phishing*, very general risks such as *loss or theft of device* and relatively new risks such as *unintentional data disclosure* through legitimate applications. Also, the experts raised a concern over more speculative risks such as *surveillance on network level* and *vendor backdoors*. Most of the identified risks are strongly dependent on the user's own actions and security awareness. Therefore, promoting smartphone security awareness among end-users should be beneficial.

Bayesian networks are found to be an effective method for documenting and analysing causal knowledge of domain experts. The model lends itself well to different types of sensitivity analysis, which would be especially useful when analysing potential controls and mitigants for risks. The expert elicitation method designed was easy for experts to understand and delivered accurate results. The process was however time-consuming, which could be eased with more effective tools. Both Bayesian networks and the expert elicitation method could be applied to other risk assessments as well. Further research is warranted for developing more effective tools and methods for expert elicitation and consolidation of results.

As is, the model can be utilized by security specialists and IT personnel to determine and communicate the most relevant risks in their environment. As future research, the Bayesian network model could be extended with controls and mitigants, which could reduce the probability of risk events or severity of consequences. With the said extension, the model could be used to analyse specific actions for strengthening the smartphone security of organisations or end-users.

References

[1] IDG Global Solutions. *IDG Global Mobile 2014 Survey*. London: IDG Global Solutions, 2014. [Online] Available from: `http://idgknowledgehub.com/mobileidg/idg-mobile-survey/` [Accessed 4 March 2015]

[2] Salesforce. *2014 Mobile Behaviour Report*. Salesforce, 2014. [Online] Available from: `https://www.exacttarget.com/2014-mobile-behavior-report` [Accessed 3 March 2015]

[3] Duggan, M., Smith, A. *Cell Internet Use 2013*. Washington: Pew Research Center, 2013. [Online] Available from: `http://pewinternet.org/Reports/2013/Cell-Internet.aspx` [Accessed 3 March 2015]

[4] Nielsen. *The Digital Consumer*. The Nielsen Company, 2014. [Online] Available from: `http://www.nielsen.com/us/en/reports/2014/the-us-digital-consumer-report.html` [Accessed 3 March 2015]

[5] comScore. *The U.S. Mobile App Report*. comScore, 2014. [Online] Available from: `http://www.smartinsights.com/mobile-marketing/mobile-marketing-analytics/mobile-marketing-statistics/` [Accessed 3 March 2015]

[6] Movable Ink. *US Consumer Device Preference Report*. Movable Ink, 2014. [Online] Available from: `http://info.movableink.com/device-report-q1-2014?utm_source=prweb` [Accessed 3 March 2015]

[7] Board of Governors of the Federal Reserve System. *Consumers and Mobile Financial Services 2014*. Washington: Board of Governors of the Federal Reserve System, 2014. [Online] Available from: `http://www.federalreserve.gov/econresdata/consumers-and-mobile-financial-services-report-201403.pdf` [Accessed 3 March 2015]

[8] Mander, J. *GWI Device Summary Q3 2014*. London: GlobalWebIndex, 2014. [Online] Available from: `http://insight.globalwebindex.net/device-q3-2014` [Accessed 3 March 2015]

[9] Peltier, T. *Information Security Risk Analysis*, 2nd Edition. Boca Raton: CRC Press, 2005.

[10] Theoharidou, M., Mylonas, A., Gritzalis, D. *A Risk Assessment Method for Smartphones*. IFIP Advances in Information and Communication Technology, 2012, volume 376, p. 443-456.

[11] Hogben, G., Dekker, M. *Smartphone security: Information Security risks, opportunities and recommendations for users* European Network and Information Security Agency (ENISA), 2010. [Online] Available from: `https://www.enisa.europa.eu/activities/identity-and-`

trust/risks-and-data-breaches/smartphones-information-security-risks-opportunities-and-recommendations-for-users [Accessed 3 March 2015]

[12] OWASP Mobile Security Project. *Top 10 Mobile Risks.* Open Web Application Security Project (OWASP), 2014. [Online] Available from: https://www.owasp.org/index.php/Projects/OWASP_Mobile_Security_Project_-_Top_Ten_Mobile_Risks [Accessed 3 March 2015]

[13] Shabtai, A., Fledel, Y., Kanonov, U., Elovici, Y, Dolev, S., Glezer, C. *Google Android: A Comprehensive Security Assessment.* IEEE Security and Privacy, 2010, volume 8 (2), p. 35—44.

[14] Milligan, P., Hutcheson, D. Business Risks and Security Assessment for Mobile Devices. In *Proceedings of the 8th Conference on 8th WSEAS Int. Conference on Mathematics and Computers in Business and Economics*, Dallas, Texas, USA, 2007, volume 8, p. 189-193.

[15] Khan, S., Nauman, M., Othman, A., Musa, S. How Secure is your Smartphone: An Analysis of Smartphone Security Mechanisms. In *International Conference on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec)*, Kuala Lumpur, Malaysia, 2012, p. 76-81.

[16] Mylonas, A., Dritsas, S., Tsoumas, B., Gritzalis, D. Smartphone security evaluation - the malware attack case. In *Proceedings of the International Conference on Security and Cryptography (SECRYPT)*, Seville, Spain, 2011, p. 25-36.

[17] Jeon, W., Kim, J., Lee, Y., Won, D. *A Practical Analysis of Smartphone Security.* Human Interface and the Management of Information, 2011, volume 6771, p. 311-320.

[18] Wang, Y., Streff, K., Raman, S. *Smartphone Security Challenges.* Computer, 2012, volume 45, p. 52-58.

[19] Enck, W., Octeau, D., McDaniel, P., Chaudhuri, S. A study of android application security. In *Proceedings of the 20th USENIX conference on Security*, Berkeley, CA, USA, 2011, p. 21.

[20] Ledermüller, T., Clarke, N.L. *Risk Assessment for Mobile Devices.* Trust, Privacy and Security in Digital Business, 2011, volume 6863, p. 210—221.

[21] Souppaya, M., Scarfone, K. *Guidelines for Managing the Security of Mobile Devices in the Enterprise.* National Institute of Standards and Technology (NIST) Special Publication 800-124, 2013. [Online] Available from: http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-124r1.pdf [Accessed 4 March 2015]

[22] Fenton, N., Neil, M. *Risk Assessment and Decision Analysis with Bayesian Networks.* Noca Raton: CRC Press, 2013.

[23] Nadkarni, S., Shenoy, P. *A causal mapping approach to constructing Bayesian networks.* Decision Support Systems, 2004, volume 38, p. 259-181.

[24] Weber, P., Medina-Oliva, G., Simon, C., Iung, B. *Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas.* Engineering Applications of Artificial Intelligence, 2012, volume 25 (4), p. 671-682.

[25] Gulvanessian, H., Holicky, M. *Determination of actions due to fire: recent developments in Bayesian risk assessment of structures under fire.* Progress in Structural Engineering and Materials, 2002, volume 3 (4), p. 346-352.

[26] Hudson, L., Ware, B., Laskey, K., Mahoney, S. *An Application of Bayesian Networks to Antiterrorism Risk Management for Military Planners*, 2002. [Online] Available from: `http://www.mathcs.emory.edu/~whalen/Papers/BNs/KathyLanskey/Antiterrorism.pdf` [Accessed 4 March 2015]

[27] Kim, M., Seong, P. *A computational method for probabilistic safety assessment of I&C systems and human operators in nuclear power plants.* Reliability Engineering & System Safety, 2006, volume 91 (5), p. 580–593.

[28] Cornalba, C., Giudici, P. *Statistical models for operational risk management.* Physica A: Statistical Mechanics and its Applications, 2004, volume 338 (1-2), p. 166-172.

[29] Russel A., Quigley J., Van der Meer R. *Modelling the reliability of search and rescue operations with Bayesian Belief Networks.* Reliability Engineering & System Safety, 2008, volume 93 (7), p. 940-949.

[30] Trucco P., Cagno E., Ruggeri F., Grande O. *A Bayesian Belief Network modelling of organisational factors en risk analysis: A case study in maritime transportation.* Reliability Engineering & System Safety, 2008, volume 93 (6), p. 845-856.

[31] Duijm, N.J. *Safety-barrier diagrams as a safety management tool.* Reliability Engineering & System Safety, 2008, volume 94 (2), p. 332-341.

[32] Røed, W., Mosleh, A., Vinnem, J.E., Aven, T. *On the Use of Hybrid Causal Logic Method in Offshore Risk Analysis.* Reliability Engineering & System Safety, 2008, volume 94 (2), p. 445-455.

[33] Hanea D., Ale B. *Risk of human fatality in building fires: A decision tool using Bayesian networks.* Fire Safety Journal, 2009, volume 44 (5), p. 704-710.

[34] Cheon S-P., Kim S., Lee S-Y., Lee, C-B. *Bayesian networks based rare event prediction with sensor data.* Knowledge-Based Systems, 2009, volume 22 (5), p. 336-343.

[35] Mo, S. Beling, P. Member, Crowther, K. Quantitative Assessment of Cyber Security Risk using Bayesian Network-based model. In *Systems and Information Engineering Design Symposium*, Charlottesville, VA, 2009, p. 183-187.

[36] Noel, S., Jajodia, S., Wang, L., Singhal, A. *Measuring Security Risk of Networks Using Attack Graphs.* International Journal of Next Generation Computing, 2010, volume 1 (1), p. 1-11.

[37] Khosravi-Farmad, M., Rezace, R., Harati, A., Bafghi, A. Network Security Risk Mitigation Using Bayesian Decision Networks. In *4th International eConference on Computer and Knowledge Engineering (ICCKE)*, Mashhad, Iran, 2014, p. 267-272.

[38] Frigault, M., Wang, L., Singhal, A., Jajodia, S. Measuring Network Security Using Dynamic Bayesian Network. In *32nd Annual IEEE International Computer Software and Applications*, Turku, Finland, 2008, p. 298-703.

[39] Dantu, R., Kolan, P. *Risk Management Using Behavior Based Bayesian Networks.* Intelligence and Security Informatics, 2005, volume 3495, p. 115-126.

[40] Sommestad, T., Ekstedt, M., Johnson, P. Cyber Security Risks Assessment with Bayesian Defense Graphs and Architectural Models. In *42nd Hawaii International Conference on System Sciences*, Big Island, HI, USA, 2009, p. 1-10.

[41] Cie, P., Li, J., Ou, X., Liu, P., Levy, R. Using Bayesian Networks for Cyber Security Analysis. In *IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, Chicago, IL, USA, 2010, 211-220.

[42] Oulasvirta, A., Rattenbury, T., Ma, L., Raita, E. *Habits make smartphone use more pervasive* Personal and Ubiquitous Computing, 2012, volume 16 (1), p. 105-114.

[43] Franko, O., Tirrell, T. *Smartphone App Use Among Medical Providers in ACGME Training Programs.* Journal of Medical Systems, 2012, volume 36 (5), p. 3135-3139.

[44] Worldwide Quarterly Mobile Phone Tracker. *Smartphone OS Market Share, Q4 2014.* International Data Corporation, 2014. [Online] Available from: `http://www.idc.com/prodserv/smartphone-os-market-share.jsp` [Accessed 4 March 2015]

[45] Faruki, P., Bharmal, A. Laxmi, V., Ganmoor, V., Gaur, M., Conti, M., Rajarajan, M. *Android Security: A Survey of Issues, Malware Penetration and Defenses.* IEEE Communications Surveys & Tutorials, 2014, volume PP (99), p. 1.

[46] Egners, A., Marschollek, B., Meyer, U. *Hackers in Your Pocket: A Survey of Smartphone Security Across Platforms*, 2012. [Online] Available from: `https://itsec.rwth-aachen.de/publications/ae_hacker_in_your_pocket.pdf` [Accessed 4 March 2015]

[47] Google, Android developers. *Sensors Overview*, [no date]. Google. [Online] Available from: `http://developer.android.com/guide/topics/sensors/sensors_overview.html` [Accessed 4 March 2015]

[48] Strategy Analytics. *Worldwide Smartphone Population Tops 1 Billion in Q3 2012*, 17 Oct 2012. [Online] Available from: `http://www.businesswire.com/news/home/20121017005479/en/Strategy-Analytics-Worldwide-Smartphone-Population-Tops-1` [Accessed 4 March 2015]

[49] Strategy Analytics. *Global Mobile Phone Shipments Reach 460 Million Units in Q3 2014*, 30 Oct 2014. [Online] Available from: `http://blogs.strategyanalytics.com/WDS/post/2014/10/30/Strategy-Analytics-Global-Mobile-Phone-Shipments-Reach-460-Million-Units-in-Q3-2014.aspx` [Accessed 4 March 2015]

[50] Omlis, *Global Mobile Payment Snapshot 2014*, 5 Aug 2014. [Online] Available from: `http://www.omlis.com/omlis-media-room/worldwide-use-of-mobile-payments/` [Accessed 4 March 2015]

[51] Rausand, M. *Risk Assessment: Theory, Methods, and Applications.* New Jersey: Wiley, 2011.

[52] Bayraktarli Y., Ulfkjaer J., Yazgan U., Faber M. On the application of bayesian probabilistic networks for earthquake risk management. In *9th International Conference on Structural Safety and Reliability (ICOSSAR 05)*, Rome, Italy, 2005.

[53] Straub D. Natural hazards risk assessment using Bayesian networks. In *9th International Conference on Structural Safety and Reliability (ICOSSAR 05)*, Rome, Italy, 2005.

[54] Eunchang, L., Park, Y., Shin, J. *Large engineering project risk management using a Bayesian belief network.* Expert Systems with Applications, 2009, volume 36 (3), p. 5880-5887.

[55] Fenton, N., Neil, M., Caballero, J. *Using Ranked Nodes to Model Qualitative Judgments in Bayesian Networks.* IEEE Transactions on Knowledge and Data Engineering, 2007, volume 19 (10), p. 1420-1432.

[56] Foss, A., Johansen, P., Hager-Thoresen, F. *Secret surveillance of Norway's leaders detected*, Aftenposten, 2014. [Online] Available from: `http://www.aftenposten.no/nyheter/iriks/Secret-surveillance-of-Norways-leaders-detected-7825278.html` [Accessed 5 April 2015]

[57] Ruggiero, P., Foote, J. *Cyber Threats to Mobile Phones.* United States Computer Emergency Readiness Team, Pittsburgh, PA, 2011. [Online] Available from: `https://www.us-cert.gov/sites/default/files/publications/cyber_threats-to_mobile_phones.pdf` [Accessed 5 April 2015]

[58] Becher, M., Freiling, F., Hoffmann, J., Holz, T., Uellenbeck, S., Wolf, C. Mobile Security Catching Up? Revealing the Nuts and Bolts of the Security of Mobile Devices. In *IEEE Symposium on Security and Privacy (SP)*, Berkeley, CA, 2011, p. 96 - 111.

[59] Vesselkov, A., Riikonen, A., Hämmäinen, H. *Mobile Handset Population in Finland 2005-2013*, Aalto University Department of Communications and Networking, 2014. [Online] Available from: `https://research.comnet.aalto.fi/public/Mobile_Handset_Population_2005-2013.pdf` [Accessed 5 April 2015]

[60] Karikoski, J., Soikkeli, T. *Contextual usage patterns in smartphone communication services.* Personal and Ubiquitous Computing, 2013, volume 17 (3), p. 491-502.

[61] Statista, *Number of smartphone users in the U.S. from 2010 to 2018 (in millions)*, 2015. [Online] Available from: `http://www.statista.com/statistics/201182/forecast-of-smartphone-users-in-the-us/` [Accessed 24 April 2015]

[62] Consumer Reports National Research Center, *3.1 Million Smart Phones Were Stolen In 2013, Nearly Double the Year Before*, Consumer Reports, 2014. [Online] Available from: `http://pressroom.consumerreports.org/pressroom/2014/04/my-entry-1.html` [Accessed 24 April 2015]

[63] Motive Security Labs, *Motive Security Labs malware report – H2 2014*, Alcatel-Lucent, 2014. [Online] Available from: `https://resources.alcatel-lucent.com/asset/184652` [Accessed 24 April 2015]

[64] Verizon Enterprise Solutions, *Verizon Data Breach Investigation Report*, Verizon, 2015. [Online] Available from: `http://www.verizonenterprise.com/DBIR/` [Accessed 24 April 2015]

[65] Boodaei, M., *Mobile Users 3 Times More Vulnerable to Phishing Attacks*, 2011. [Online] Available from: `http://securityintelligence.com/mobile-users-3-times-more-vulnerable-to-phishing-attacks/` [Accessed 20 June 2015]

[66] Klein, A. *The Golden Hour of Phishing Emails*, 2010. [Online] Available from: `http://securityintelligence.com/the-golden-hour-of-phishing-emails/` [Accessed 20 June 2015]

[67] Huang, K., Henrion, M. Efficient Search-Based Inference for Noisy-OR BeliefNetworks. In *Twelfth Conference on Uncertainty in Artificial Intelligence*, Portland, OR, 1996, 325-331.

[68] Díez, F.J. Parameter adjustment in Bayes networks: the generalized noisy or-gate. In *Ninth Conference on Uncertainty in Artificial Intelligence*, Washington D.C, 1993, 99-105.

[69] Ponemon Institute, *The Lost Smartphone Problem*, Ponemon Institute and McAfee, 2011. [Online] Available from: `http://pdf.thepdfportal.net/PDFFiles/91011.pdf` [Accessed 23 June 2015]

[70] Monrose, F. White, A. Raguram, R., Frahm, J-M., Goswami, D. *iSpy: Using Reflections To Spy On iPhones*, 2011. [Online] Available from: `https://packetstormsecurity.com/files/106678/CCS2011.pdf` [Accessed 23.06.2015]

[71] FireEye, *A Comprehensive Mobile Threat Assessment of 7 Millions iOS and Android Apps*, 2015. [Online] Available from: `https://www2.fireeye.com/WEB-2015RPTMobileThreatAssessment.html` [Accessed 29.06.2015]

[72] Nohl, K., Lehl, J. *BadUSB–On accessories that turn evil*, Black Hat USA, 2014. [Online] Available from: `https://srlabs.de/blog/wp-content/uploads/2014/07/SRLabs-BadUSB-BlackHat-v1.pdf` [Accessed 29.06.2015]

[73] Felt, A., Wagner, D. *Phishing on Mobile Devices*, Workshop on Web Security and Privacy (W2SP), 2011. [Online] Available from: `http://w2spconf.com/2011/papers/felt-mobilephishing.pdf` [Accessed 1.7.2015]

[74] Barkan, E., Biham, E., Keller, N. *Instant Ciphertext-Only cryptanalysis of GSM Encrypted Communication.* Journal of Cryptology, 2008, volume 21 (3), p. 392-429.

[75] Google Report. *Android Security 2014 Year in Review*, 2015. [Online] Available from: `https://static.googleusercontent.com/media/source.android.com/fi//devices/tech/security/reports/Google_Android_Security_2014_Report_Final.pdf` [Accessed 2.7.2015]

[76] Peltola, M., Kekolahti, P. Risk Assessment of Public Safety and Security Mobile Service. In *International Conference on Availability, Reliability and Security ("ARES")*, Toulouse, France, 2015.

[77] Wang, J., Guo, M. Vulnerability Categorization Using Bayesian Networks. In *Proceedings of the Sixth Annual Workshop on Cyber Security and Information Intelligence Research*, Oak Ridge, Tennessee, USA, 2010, no. 29, p. 1-4.

[78] Fischhoff, B., Slovic, P., Lichtenstein, S. *Fault trees: Sensitivity of estimated failure probabilities to problem representation.* Journal of Experimental Psychology: Human Perception and Performance, 1978, volume 4(2), p. 330-344.

[79] Kemeny, J.G., Snell, J.L. *Finite markov chains.* Princeton, NJ: van Nostrand, 1960.

[80] Murata, T. *Petri nets: Properties, analysis and applications.* Proceedings of the IEEE, 1989, volume 77(4), p. 541-580.

[81] Box, G.E., Tiao, G.C. *Bayesian inference in statistical analysis.* John Wiley & Sons, 1973.

[82] Uusitalo, L. *Advantages and challenges of Bayesian networks in environmental modelling.* Ecological Modelling, 2007, volume 203(3-4), p. 312-318.

[83] Symantec, *The Symantec Smartphone Honey Stick Project*, 2012. [Online] Available from: `http://www.symantec.com/content/en/us/about/presskits/b-symantec-smartphone-honey-stick-project.en-us.pdf` [Accessed 17.7.2015]

[84] Singh, M., Valtorta, M. *Construction of Bayesian network structures from data.* International Journal of Approximate Reasoning, 1993, volume 12(2), p. 111-131. [Online] Available from: `http://www.sciencedirect.com/science/article/pii/0888613X9400016V` [Accessed 17.7.2015]

[85] Chickering, D.M., Heckerman, D., Meek, C. A Bayesian approach to learning Bayesian networks with local structure. In *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*, Providence, Rhode Island, 1997, p. 80-89. [Online] Available from: `http://arxiv.org/ftp/arxiv/papers/1302/1302.1528.pdf` [Accessed 17.7.2015]

[86] Kekolahti, P., Using Bayesian Belief Networks for modelling of Communication Service Provider Businesses. In *Proceedings of the 8th Bayesian modelling Applications Workshop*, Barcelona, 2011.

[87] BayesiaLab, Bayesian network publishing and automatic learning program, 2010, release 5.0. [Online] Available from: `http://www.bayesia.com/en/products/bayesialab.php` [Accessed 19.7.2015]

[88] Bayes Server, Machine learning and reasoning software using Bayesian networks, 2015, release 6.12. [Online] Available from: `http://www.bayesserver.com/` [Accessed 19.7.2015]

[89] Netica, A complete software package to solve problems using Bayesian Belief Networks and influence diagrams, 2015, release 5.15. [Online] Available from: `https://www.norsys.com/` [Accessed 19.7.2015]

[90] AgenaRisk, Bayesian Network and Simulation Software for Risk Analysis and Decision Support, 2015, release 6.1. [Online] Available from: `http://www.agenarisk.com/` [Accessed 19.7.2015]

[91] Rissanen, J. *Modeling by shortest data description.* Automatica, 1978, volume 14(5), p. 465-471.

[92] Aldrich, J. *R.A. Fisher and the making of maximum likelihood 1912-1922.* Statistical Science, 1997, volume 12(3), p. 162-176. [Online] Available from: `http://projecteuclid.org/euclid.ss/1030037906` [Accessed 30.7.2015]

[93] Tversky, A., Kahneman, D. *Judgment under Uncertainty: Heuristics and Biases.* Science, 1974, volume 185(4157), p. 1124-1131.

[94] Hänninen, M., Kujala, P. *Influences of variables on ship collision probability in a Bayesian belief network model. Reliability Engineering & System Safety*, 2012, volume 102, p. 27-40.

[95] Helle, I., Lecklin, T., Jolma, A., Kuikka, S. *Modeling the effectiveness of oil combating from an ecological perspective – A Bayesian network for the Gulf of Finland; the Baltic Sea.* Journal of Hazardous Materials, 2011, volume 185(1), p. 182-192.

Appendix A Examples of consequence severity distributions

Example consequence severity distributions were shown to the experts during the interviews to ease the process of describing a severity distribution. An example of these distributions is shown in figure A1.

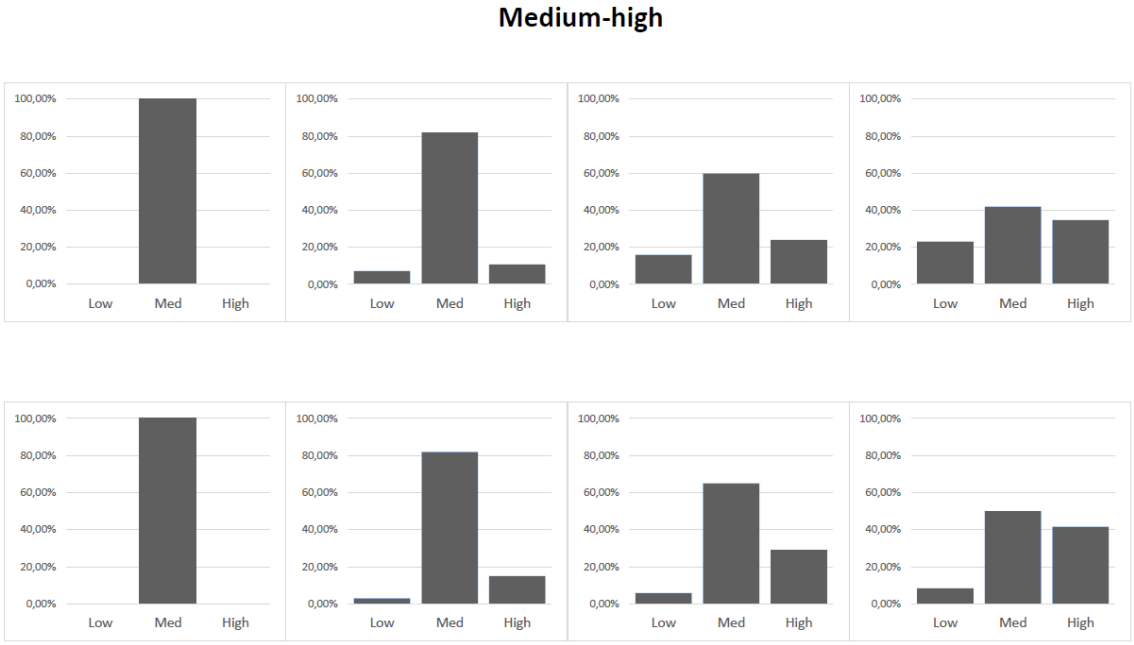
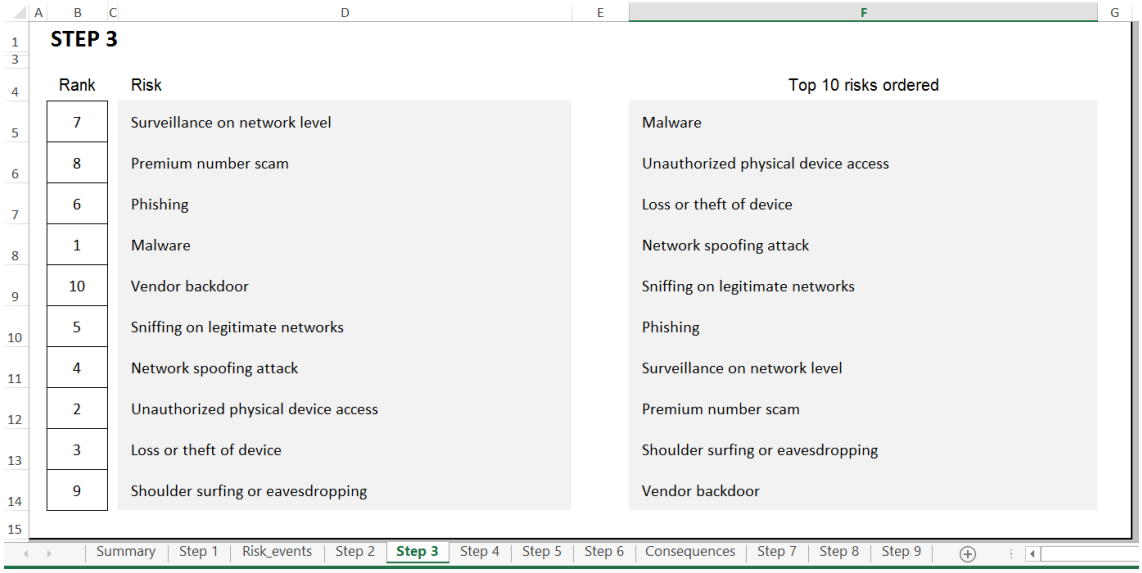


Figure A1: Example consequence severity distributions

Appendix B Tool for facilitating expert interviews

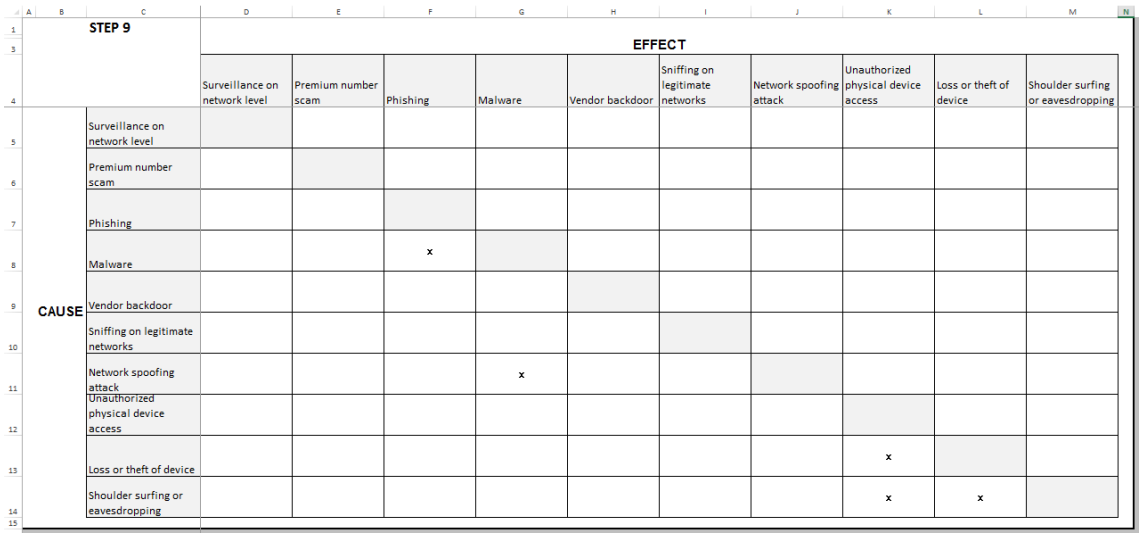
An Excel-based tool was designed in order to facilitate the stage 1 interviews with experts such as described in section 3.3.2. Figures B1 and B2 show two views of this tool, one for ordering the identified risks by importance and another for identifying the existence of causal relationships between these risks.



STEP 3

Rank	Risk	Top 10 risks ordered
7	Surveillance on network level	Malware
8	Premium number scam	Unauthorized physical device access
6	Phishing	Loss or theft of device
1	Malware	Network spoofing attack
10	Vendor backdoor	Sniffing on legitimate networks
5	Sniffing on legitimate networks	Phishing
4	Network spoofing attack	Surveillance on network level
2	Unauthorized physical device access	Premium number scam
3	Loss or theft of device	Shoulder surfing or eavesdropping
9	Shoulder surfing or eavesdropping	Vendor backdoor

Figure B1: Step 3 of Excel-based tool for facilitating expert interviews



STEP 9

CAUSE \ EFFECT	Surveillance on network level	Premium number scam	Phishing	Malware	Vendor backdoor	Sniffing on legitimate networks	Network spoofing attack	Unauthorized physical device access	Loss or theft of device	Shoulder surfing or eavesdropping
Surveillance on network level										
Premium number scam										
Phishing										
Malware			x							
Vendor backdoor										
Sniffing on legitimate networks										
Network spoofing attack				x						
Unauthorized physical device access										
Loss or theft of device								x		
Shoulder surfing or eavesdropping								x	x	

Figure B2: Step 9 of Excel-based tool for facilitating expert interviews

Appendix C Tool for calculating NPTs

The node probability tables (NPT) in the Bayesian network were calculated using an Excel-based tool. Figure C1 shows a part of the consolidated data from existing research and expert interviews while figure C2 shows a part of a final NPT, which was then exported to AgenaRisk.

	A	B	C	D	E	F
1	**CONSEQUENCE:**	**Severity distribution**				
2	**Risk event**	**Negligible**	**Low**	**Medium**	**High**	**Share/total**
3	**Latent threats**	**96,18 %**	**1,41 %**	**1,46 %**	**0,94 %**	**7,00 %**
4	**Unintentional data disclosure**	**92,95 %**	**4,56 %**	**1,95 %**	**0,55 %**	**43,93 %**
5	**Malware**	**61,19 %**	**18,79 %**	**13,68 %**	**6,34 %**	**0,42 %**
6	**Unauthorized physical device access**	**58,33 %**	**19,31 %**	**16,95 %**	**5,41 %**	**1,34 %**
7	**Network spoofing attack**	**81,50 %**	**12,83 %**	**4,12 %**	**1,55 %**	**1,81 %**
8	**Backdoor (vendor)**	**85,83 %**	**2,45 %**	**6,02 %**	**5,69 %**	**0,77 %**
9	**Surveillance on network level**	**79,00 %**	**9,32 %**	**6,79 %**	**4,89 %**	**0,26 %**
10	**Shoulder surfing or eavesdropping**	**90,33 %**	**5,44 %**	**3,22 %**	**1,01 %**	**4,20 %**
11	**Phishing**	**78,86 %**	**7,74 %**	**8,61 %**	**4,79 %**	**0,51 %**
12	**Sniffing on legitimate networks**	**85,72 %**	**9,32 %**	**3,60 %**	**1,36 %**	**1,33 %**
13						

Figure C1: Partial consolidated data regarding the consequence *leakage of confidential data*

H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
NPT																
Negligible	0,96	0,89	0,59	0,55	0,56	0,52	0,34	0,32	0,78	0,73	0,48	0,45	0,46	0,43	0,28	0,26
Low	0,04	0,08	0,21	0,23	0,22	0,24	0,28	0,29	0,16	0,19	0,27	0,29	0,28	0,29	0,31	0,31
Medium	0,00	0,02	0,14	0,15	0,17	0,18	0,27	0,28	0,04	0,06	0,17	0,18	0,20	0,21	0,29	0,30
High	0,00	0,01	0,06	0,07	0,05	0,06	0,11	0,12	0,02	0,02	0,08	0,08	0,07	0,07	0,13	0,13
Unintenti	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FA
Malware	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FA
Unauthori	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FA
Network s	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FA
Backdoor	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TI
Surveillan	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FA
Shoulder	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FA
Phishing	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FA
Sniffing o	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FA

Figure C2: Partial NPT for the consequence *leakage of confidential data*

Appendix D Risk event and consequence descriptions used in stage 2 interviews

Risk events	Explanations
Network device spoofing attack	*User connects to a rogue network device that impersonates a trusted device such as a WLAN access point, cellular base station or Bluetooth device. Depending on the technology, connecting might happen automatically or require action from the user. By definition, the rogue network device has been deployed in order to monitor and/or intercept data from users.*
Sniffing on legitimate data network	*Packet sniffing performed by an attacker on a data network used by the smartphone, most likely an unencrypted WLAN network. Can also lead to a Man in the Middle -attack.*
Surveillance on cellular network	*Surveillance or tracking performed on network level, independent of a user's device's security. This could be performed for example by accessing the SS7 signaling network, a local operator's core network, a base station or data collected by a network operator.*
Malware	*Any malicious software. Can be for example disguised as a legitimate application or embedded into a legitimate application. Includes all kinds of spyware, ransomware, diallerware, surveillance software etc.*
Premium number fraud	*Billing caused by calls or messages to premium rate services without the user's approval and consent. Includes both tricking a user into making the call/message themselves as well as malware that performs the action without a user's knowledge.*
Vendor backdoor	*A backdoor or information collection agent collects data from or in some way controls the device. By definition, the backdoor was present when the device was bought as opposed to malware which infects devices in use. Backdoor can be for example enforced and used by a government.*
Loss or theft of device	*Device gets stolen or is lost.*
Unauthorized physical device access	*Accessing the device locally without the owner's or user's permission.* ***Note:*** *Unauthorized access by family members is excluded.*
Technical failure of device	*Any technical failure of hardware or software (crashing) that meaningfully affects the device user.*
Technical failure of network	*Cellular network failure that significantly affects smartphone use for at least 1 hour.*
Unintentional data disclosure	*Unknowingly disclosing data through a legitimate application for example due to incomplete understanding of privacy policy/settings or due to settings changes caused by another user/application.*
Phishing	*Misleading a user to willingly give out sensitive information such as user credentials.*
Shoulder surfing or eavesdropping	*Purposefully watching and/or recording video of a user use his smartphone. Purposefully listening and/or recording audio of a user speaking into his smartphone.* ***Note:*** *Shoulder surfing or eavesdropping by family members is excluded.*

Consequence	Explanation	Examples
Leakage of personal data	*Any personal data leaks to an unauthorized party or an unauthorized party is able to impersonate the user in another service such as Facebook, a mobile banking application or the cellular network. Impersonation could be possible for example due to auto-login settings, credential discovery or MitM attacks.* *Personal data includes for example the following types of data.* • *Data stored on smartphone such as contacts, photos, messages and documents* • *Usage information and sensor data such as location data* • *Communication data such as messages, phone calls or other data transfers* • *Credentials to other services* • *Credit card and other personal identification information*	Low impact: • Location data • Unspecific usage information Medium impact: • Photos, messages and documents, unless especially sensitive • The contents of a single phone call, unless especially sensitive • Impersonating user in a non-sensitive service High impact: • Credit card or personal identification information • Credentials to other services • Systematic surveillance of all phone calls • Impersonating user in a sensitive service
Leakage of confidential data (business/government)	*Any confidential data related to a user's employer or other associated business or governmental agency leaks to an unauthorized party. The assumption here is that each user is employed and could potentially use their smartphone for storing, accessing or communicating confidential information. This data can include for example:* • *Data stored on smartphone such as emails and documents* • *Communication data such as messages and phone calls* • *In some cases also usage and sensor data*	Low impact: • Leakage of inconsequential but non-public information Medium impact: • Leakage of sensitive but non-critical information • Minor effect on company or government functions High impact: • Leakage of trade or governmental secrets • Leakage of data that affects stock market • Disrupting company or government functions
Data loss or corruption	*Any data with value is lost or corrupted. This consequence includes situations where it would be technically possible to recover the data, but where a typical user does not have knowledge or resources to perform the recovery and thus they do not have access to the data anymore.*	Low impact: • Loss of recoverable or non-important data Medium impact: • Loss of important but non-critical data High impact: • Extensive data loss or loss of critical data
Unavailability of device or services	*A situation where the device or a service normally used by the user is partially or completely unavailable. For example:* • *Smartphone does not turn on* • *Cellular network is unavailable* • *Cellular network is available but data connections do not work*	Low impact: • Inconvenience to the user Medium impact: • Inability to perform necessary actions High impact: • Unavailability of a service for long periods of time • Unavailability of emergency or other critical services
Financial consequences	*Direct financial consequences to the user or owner of the device. For example excessive billing, repair expenses or the cost of purchasing a new device.*	Low impact: • 0 € < X < 20 € Medium impact: • 20 € ≤ X < 200 € High impact: • X ≥ 200 €