Strengthening Mobile Network Security Using Machine Learning

Van Thuan Do[1], Paal Engelstad[2], Boning Feng[2], and Thanh van Do[3,4(✉)]

[1] Linus AS, Martin Linges vei 15, 1364 Fornebu, Norway
t.do@linus.no
[2] College of Applied Sciences, Oslo and Akershus University, Pilestredet 46, 0167 Oslo, Norway
{paal.engelstad,boning.feng}@hioa.no
[3] Telenor ASA, Snarøyveien 30, 1331 Fornebu, Norway
thanh-van.do@telenor.com
[4] Norwegian University of Science and Technology, 7031 Trondheim, Norway

Abstract. Lately, several episodes of tapping and tracking of mobile phones in Europe including Norway have been revealed, showing the vulnerabilities of both the mobile network and mobile phones. A better protection of the user's confidentiality and privacy is urgently required. This paper will present an innovative mobile network security system using machine learning. The paper will start with a vulnerability and threat analysis of the evolving mobile network, which is a fusion of mobile wireless technologies and Internet technologies, complemented with the Internet of Things. The main part of the paper will concentrate on clarifying how machine learning can help improving mobile network security. The focus will be on elucidating what makes machine learning superior to other techniques. A special case study on the detection of IMSI Catcher, the fake base station that is used in mobile phone tracking and tapping, will be explained.

Keywords: Mobile network security · Mobile privacy · Cyber security · Cyber attacks

1 Introduction

Until lately the mobile network was perceived as quite secured compared to the Internet since the users benefit of stronger encryption provided by the SIM card [1]. Unfortunately, the recent phone tapping incidents revealed by the former CIA agent Snowden have shocked the whole world and raised doubt about the security of the mobile network [2]. In fact, the reduced prices of hardware equipment resulting from advances in microelectronics combined with the availability of open source mobile communication software have made the attacks on mobile networks both easier and more affordable. The need for better protection of user security and privacy is more urgent than never. The biggest challenge is, however, due to the fact that the mobile network resulting from a fusion of mobile and Internet technologies, inherits the weaknesses of both parties and worst suffers also of the unknown ones born by the marriage.

© Springer International Publishing Switzerland 2016
M. Younas et al. (Eds.): MobiWIS 2016, LNCS 9847, pp. 173–183, 2016.
DOI: 10.1007/978-3-319-44215-0_14

This paper will present an innovative mobile network security system using machine learning. The paper will start with an overview of the vulnerabilities and threats of the modern mobile network, which is a fusion of mobile wireless technologies and Internet technologies, complemented with the Internet of Things. The main part of the paper will concentrate on clarifying how machine learning can help improving the mobile network security. The focus will be on elucidating what makes machine learning superior to other techniques. A special case study on the detection of IMSI Catcher, the fake base station that is used in mobile phone tracking and tapping, will be explained. A machine learning based IMSI catcher detection is described. The paper ends with suggestions of future works in the area of Machine Learning and mobile network security.

2 Related Works

Mobile network security has attracted more attention lately but the research activities are still limited to the ones of a few communities that will be briefly described in the coming sections.

2.1 Security Research Labs (SRLabs)

The SRLabs [3] in Berlin led by the famous German Cryptographer and security researcher Karsten Nohl has considerable activities related to detection of mobile phone tapping. SRLabs has a collection of tools for the assessment of mobile network security.

2.2 P1 Security

P1 Security (Priority One Security) [4] is a company led by Philippe Langlois, a well-known security expert, which is dedicated to providing top security products and services for high-expertise security areas. P1 Security has a Telecom Security Task Force, which is a research think tank and consulting network in Telecom sector.

2.3 SBA Research

SBA Research [5] is an Austrian research center for Information Security funded by the national initiative for COMET Competence Centers for Excellent Technologies and consisting of 25 companies, 4 Austrian universities and several international research partners. The center is focusing on challenges ranging from organizational to technical security and has recently a few activities on mobile network security including the implementation of the IMSI catcher.

All the mentioned communities do have activities on mobile network security yielding valuable results, which are used in our research. However, none of them proposes to use machine learning in the protection of the mobile network.

3 Threats and Vulnerabilities in the Mobile Network

Although compared to the Internet the mobile network is still much more secured it is now more exposed. Designed and built as a walled garden system the mobile network has evolved to become an open system, which is in nature much more vulnerable to attacks. Further, a lot of changes have been happening since the introduction of mobile communication and reshaping the landscape dramatically. First, deregulation was removing the monopoly of telecom operators at the same time as opening the market for less trusted parties. Next, in order to pave the way for innovative services in addition to voice and short message, mobile operators have to adopt IP technology, which brought with it a series of weaknesses. Third, advances in microelectronic have made possible the production of lower price hardware equipment that could be used in the attacks against mobile networks. Last but not least, the emergence of open source mobile communication software such openBTS [6], openBSC [7], Open Source GSM Baseband software, etc. has enabled the construction of base station of a few hundred dollars, which can be used as fake base station, aka IMSI catcher [8–10] to impersonate the users.

Figure 1 shows the vulnerable entry points where attacks have been launched against the mobile network as follows:

- **Mobile phone:** mobile phones are exposed to viruses and could be crashed by attacks such as SMS of Death. OsmocomBB [11] a free Open Source GSM Baseband software implementation can be used to build hostile phones that are used in the attacks against subscribers and mobile networks.

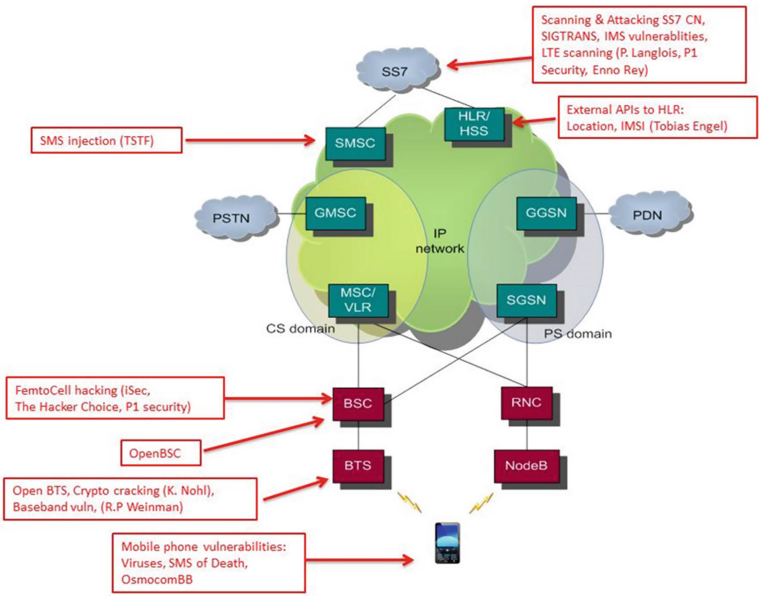

Fig. 1. Vulnerabilities in the mobile network

- **BTS (Base Transceiver Station):** Open source openBTS can be used to build fake base station aka IMSI catcher. Crypto cracking (by Karsten Nohl), Baseband vulnerabilities (by Weinman) show that the encrypted radio channel between the BTS and the mobile phone could be cracked.
- **BSC (Base Station Controller):** Fake BSC can be built using OpenBSC. Femtocell hacking can be used to penetrate the mobile core network.
- **SMSC (Short Message Service Center):** SMS injection can be used to attack SMSC.
- **HLR (Home Location Register):** Location tracking, IMSI capture (by Tobia Engel) can be launched using the HLR's external APIs
- **SS7 (Signalling Signal 7):** Scanning & attacking SS7 CN, SIGTRANS, IMS vulnerabilities, LTE scanning (by Langlois) show the vulnerabilities of SS7 networks.

4 Briefly About Machine Learning

Before examining how machine learning can contribute to securing mobile network it is worth to revise the definition of machine learning. According to Mitchell [12]:

> *"The field of machine learning is concerned with the question of how to construct computer programs that automatically improve with experience"*

He provides also a short formalism as follows:

> *"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."*

The machine's ability to learn and improve its solutions to problems is hence central in machine learning and there are different learning ways that can be chosen for the machine depending on the nature of the application as follows:

- **Supervised learning:** the machine is trained using labeled examples, such as an input where the desired output is known. For example, a piece of log file could have data points labeled either "A" (attack) or "B" (benign).
- **Unsupervised learning:** this method is used against data that has no historical labels. The system is not told the "right answer." The algorithm must figure out what is being shown. The goal is to explore the data and find some structure within them.
- **Semi-supervised learning:** this method is used for the same applications as the ones for supervised learning. But both labeled and unlabeled data for training are used – typically a small amount of labeled data together with a large amount of unlabeled data (because unlabeled data is less expensive and takes less effort to acquire).
- **Reinforcement learning:** this method discovers through trial and error which actions yield the greatest rewards.

Two of the most widely adopted machine learning methods are supervised learning with around 70 percent and unsupervised learning with 10 to 20 percent.

5 How Can Machine Learning Improve Mobile Network Security

5.1 Zero Day Attacks

As stated previously, the mobile network has considerable vulnerabilities but the worst is that not all the vulnerabilities are identified and there are probably other unknown vulnerabilities.

In addition to attacks that exploit older, more commonly known vulnerabilities that have not yet been patched, or make use of basic poor security practices, there are **Zero day attacks**.

Zero day attacks are attacks on zero day vulnerabilities, i.e. vulnerabilities that become publicly known (zero-day) on the same day of the attack or more generally, vulnerabilities which have not been patched or made public. Once known the vulnerability is not called zero day anymore but known vulnerabilities and a race for protection solutions begins.

A zero-day attack starts when a flaw or software or hardware vulnerability is exploited and attackers release malware before a developer has an opportunity to create a patch to fix the vulnerability — hence "zero-day."

Vulnerabilities may be discovered by hackers, by security companies or researchers, by the software or hardware suppliers themselves or by users. If discovered by hackers, the vulnerabilities will be exploited and kept secret for as long as possible. Hackers will circulate only through the ranks of hackers, until software or security companies become aware of it or of the attacks targeting it. These types of attacks are defined by some as 'less than zero-day' attacks.

Lately, although not yet common, attacks exploiting multiple zero day vulnerabilities has emerged. Further, in order to extend the attack window the attackers modify their tools just enough to evade detection a little bit longer. Polymorphism and metamorphism are obfuscation techniques that are used to evade detection.

Consequently, zero day attacks and their mutations are quite difficult to parry and this is where Machine Learning can come to help. It can help building a variety of profiles such as user profile, network traffic, service usage, access activities, etc. which define normal situation or normal behavior. Any deviation indicates anomalies that can trigger an alarm resulting to intervention of experts.

5.2 Challenges in the Construction of Conclusive Attack Signatures

To detect known attacks on their networks mobile operators could use signature based IDS (Intrusion Detection System). There exist several definitions of attack signature which are slightly different depending on the focus and detail level. In this paper, an attack signature is defined as a characteristic or distinctive pattern that can be searched for or that can be used in matching to previously identified attacks [13]. Unfortunately, some serious known attacks on the mobile networks do not have sufficiently distinctive patterns that can be used to distinguish them from regular benign actions. The usage of such patterns in the detection of abuses in the mobile network leads to both unacceptably high false positive rate and high false negative rate i.e. while the usage of

insufficiently distinctive patterns trigger a lot of false alarms, real abuses can still go unobserved. This situation will be illustrated with the case study IMSI catcher in the next section.

6 Case Study: The IMSI Catcher

6.1 Short About IMSI Catcher

According to [14] An IMSI catcher is a device for intercepting GSM mobile phones. It subjects the phones in its vicinity to a Man-In-The-Middle (MITM) attack by pretending to be the preferred base station in terms of signal strength. Actually, it is a fake base station that lures mobile phone to attach to it instead of legal ones. As its name tells, the IMSI catcher logs the IMSI numbers of all the mobile phones in the area, as they attempt to attach to the base station, and can determine the phone number of each individual phone. It also allows forcing the mobile phone connected to it to revert to A5/0 for call encryption (in other words, no encryption at all), making the call data easy to intercept and convert to audio. The phone calls can hence be tapped and recorded by the IMSI Catcher.

6.2 Challenges in the Detection of IMSI Catcher

IMSI catchers constitute undoubtedly a substantial threat to the users since they can be used in the invasion of the user's privacy i.e. they can be used in the surveillance of the users, in the interception of calls and short messages. However, they do not cause any harm to the mobile network neither in terms of availability and performance nor revenues and reputation. Further, they do not leave real trace since they do actually not intrude into the mobile network. This is probably the reason for that most of the IMSI catcher detection solutions are device based, i.e. a dedicated handheld device such as a GSMK CRYPTOPHONE [15] or an app like to be installed on a smartphone like Android IMSI-Catcher Detector (AIMSICD) [8], Snoopsnitch [16]. In fact, according to our knowledge there is currently no IMSI catcher detection solution that is on the network and operated by mobile operator for the protection of the users. In this case study, we are proposing a network-based IMSI catcher which is using Machine Learning to detect the presence of IMSI catcher.

6.3 Challenges in the Establishment of IMSI Catcher Signature

To detect the presence of IMSI catcher in the mobile network it is necessary to have a signature which does not exist since no detection system exists in the mobile network. Therefore, we have to build a signature by composing several indicators i.e. characteristics that could together indicate a probable presence of an IMSI catcher. A thorough study of the mobile network architecture, network elements and network interfaces has been carried out to identify the relevant indicators that are successively described in the coming sections.

6.3.1 Handover from 3G to 2G

The current mobile network is usually a composition of 2G (GSM) and 3G (UMTS) mobile networks in order to provide ubiquitous mobile coverage and support of all types of mobile phones. However, most of current mobile phones in use support both 2G, 3G and a series of frequency bands. When the mobile phones are moving around handover i.e. shift between cells of same technology, i.e. 2G-2G or 3G-3G or between cells of different technologies, i.e. 2G-3G or 3G-2G can be executed depending on:

- Signal strength variation
- Signal quality
- Load balancing between cells
- Distance between cells

Indicator HO1: Abnormal high handover from 3G to 2G in an area where 3G coverage is good could indicate the presence of an IMSI catcher jamming 3G signal and forcing mobile phones to downgrade to 2G for call tapping.

- *Dependency:* Must be used together with other indicators
- *Limitation:* 3G-2G handover is quite usual because 2G usually has better coverage than 3G
- *Quality:* High level of false positives
- *Confidence:* Low

Indicator HO2: Changes in the 2G-2G handover patterns could also indicate the arrival of an IMSI catcher.

- *Dependency:* Must be used together with other indicators
- *Limitation:* Influenced by changes in the network such as network optimalisation, errors in neighbor cells, traffic variations, atmospheric variation, etc.
- *Quality:* Very high level of false positives because 2G-2G can occur due circumstances mentioned in limitation
- *Confidence:* Low

Indicator HO3: Increase in unsuccessful handovers could also be due to the signal jamming of an IMSI catcher.

- *Dependency:* Must be used together with other indicators
- *Limitation:* Influenced by changes in the network such as network optimalisation, errors in neighbor cells, traffic variations, atmospheric variation, etc.
- *Quality:* Very high level of false positives because unsuccessful handovers can occur due circumstances mentioned in limitation
- *Confidence:* Low

6.3.2 Location Update

In the mobile networks, cells are grouped into *Location Area* (LA) for 2G networks, *Routing Area* (RA) for 3G networks. These areas are identified by an area code such as

Location Area Code (LAC), Routing Location Code (RAC). When a mobile phone moves and changes location areas it will perform a location update to inform the mobile network about the new location area such that calls can be delivered to the mobile station. A mobile terminal can change areas between 2G and 3G network without changing geographical locations. Changes in LAC or RAC can be used in the detection of IMSI catcher because they can show abnormal location areas or abnormal location update patterns due to the presence of IMSI catcher.

Indicator LU1: Increase in LAC updating can be a sign of an IMSI catcher

- *Dependency:* Must be used together with other indicators
- *Limitation:* Influenced by changes in the network such as network optimalisation, errors in neighbor cells, traffic variations, etc.
- *Quality:* High level of false positives because the increase in LAC updating can be due to circumstances stated in limitation
- *Confidence:* Medium

Indicator LU2: Sequence of last visited LAC can also indicate the presence of an IMSI catcher

- *Dependency:* Must be used together with other indicators
- *Limitation:* If the IMSI catcher is configured as a 3G base station covering the same area as the 2G cell this indicator will not be able to detect it.
- *Quality:* High level of false positives because the mobile phones can be switched off or run out of battery
- *Confidence:* Medium

6.3.3 Relation Between IMSI and IMEI

Every mobile subscriber gets assigned from her mobile operator a universal unique identity called IMSI (International Mobile Subscriber Identity), which, installed in the SIM card is used in the identification of the subscriber at connection to the mobile network. The mobile phone itself has also a unique identity called IMEI (International Mobile Equipment Identity). Normally the relation IMSI-IMEI is quite stable and recorded in mobile network. A change of the IMSI-IMEI relation can be used in the detection of IMSI catcher.

Indicator II1: Multiple IMSI One IMEI can indicate that one IMSI catcher is impersonating multiple subscribers

- *Dependency:* Relies on the database storing the relation IMSI-IMEI
- *Limitation:* This indicator can be used to detect only active IMSI catchers that are monitoring calls or intercepting SMS but fails to detect passive IMSI catchers that track the location of the user. More advanced IMSI catchers can also clone the IMEI of their target and remain invisible. Of course, this indicator fails totally when a large number of phones does not have IMEI or use the same fake IMEI as it is the case of many developing countries.

- *Quality:* medium quality to detect less advanced IMSI catchers but low quality for advanced IMSI catchers
- *Confidence:* Medium

Indicator II2: One IMSI Multiple IMEI can indicate that a particular subscriber is exposed for attack by one or more IMSI catchers.

- *Dependency:* Relies on the database storing the relation IMSI-IMEI
- *Limitation:* This indicator has the same limitations as indicator II1. In addition, the user may also have multiple devices and move her SIM card between them.
- *Quality:* medium quality to detect less advanced IMSI catchers but low quality for advanced IMSI catchers
- *Confidence:* Medium

By examining all the indicators we can conclude that none of the indicators can be used alone by itself and more importantly, by using all the indicators together it is still not sufficient to determine the presence of an IMSI catcher. Therefore, a signature based IMSI catcher detection is proved to be insufficiently efficient.

6.4 A Machine Learning Based IMSI Catcher Detection

Although the indicators described in the previous sections are not sufficient to compose a signature usable in the detection of IMSI catchers they can be used in a Machine Learning based IMSI catcher detection.

As shown in Fig. 2 the online-detection part contains different anomaly detectors, each of which uses an indicator i.e. HO1, HO2, HO3, LU1, LU2, II1 and II2 to define normal and abnormal behavior. A simplest form of the ensemble model is the majority voting between the different detectors but a weighted voting may also be considered in later phases. Several machine-learning algorithms, such as one-class Support Vector Machines [17] and Neural Networks, can be used as anomaly detectors.

Following the suggestions from the ensemble detector, security experts would then look at suspicious places to verify if there any true IMSI catchers at a point in time. The feedback from the security experts is then given back to off-line learning part to update the models where the normal behavior was defined.

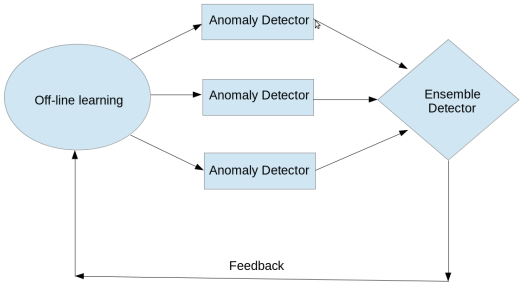

Fig. 2. The proposed Machine Learning based IMSI catcher detection

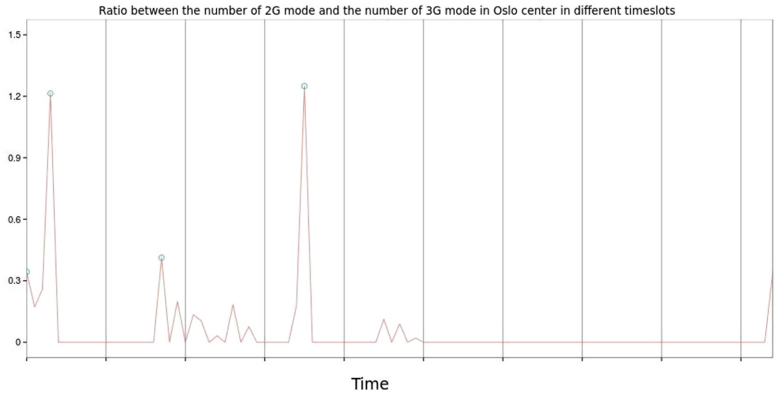

Fig. 3. Experiments on 2G/3G modes with Aftenposten data set

At the present stage of our project no real data set has been yet collected from the mobile network and for illustration sake we did some experiments on the public available data set related to IMSI catcher detection from Aftenposten [18]. The data came from a handset while interacting with the mobile network and possibly with the IMSI catchers as well, and we cannot show the advantages of machine-learning in correlating different events from many different devices. However, the main objective of this experiment is to show that there is a potential of applying machine learning techniques to facilitate the detection process.

For our simple experiment, from the data set we were interested in the frequency of the mode change between 2G and 3G. Our hypothesis is that the high value of the frequency would indicate abnormality in an area.

We split the data by equal time slots and calculate the ratio between the number of 2G and 3G in each time slot. We applied the anomaly detection algorithm named S-H-ESD from Twitter [19] to detect abnormalities for those obtained ratio values. The result is shown in Fig. 3. The three high spikes, which denote the abnormalities, indicate the possible presence of IMSI catcher.

In the next stage, we will deploy an IMSI catcher that we are building using openBTS in a test mobile network. We will collect data and feed them into the anomaly detectors. The results from the Ensemble Detector will be used to train the off-line learning.

7 Conclusion

In this paper we explain the vulnerabilities of the mobile networks which can exploited by malicious attacks. Unfortunately, current Intrusion Detection Systems using signature are not sufficiently efficient for mobile networks. This is because of zero-day vulnerabilities and attacks without distinctive signature. To remedy the situation, we propose to adopt machine learning technique that makes use of indicators to define normal and abnormal situation. Such detection could be much more efficient than

existing solutions. The paper presents an IMSI catcher detection as a case study and a pilot Machine Learning based IMSI catcher Detection is elaborated for verification. As further, we will establish a test mobile network and an IMSI catcher by using openBTS such that experiments can be carried out and data collected for the Machine Learning detection.

References

1. van Thanhe, D., Jørstad, I., van Thuan, D.: Strong authentication for web services with mobile universal identity. In: Younas, M., Awan, I., Mecella, M. (eds.) MobiWIS 2015. LNCS, vol. 9228, pp. 27–36. Springer, Heidelberg (2015). ISSN 0302-9743, ISSN 1611-3349 (electronic), ISBN 978-3-319-23143-3, ISBN 978-3-319-23144-0 (e-book)
2. Foss, A.B., Johansen, P.A., Hager-Thoresen, F.: Secret surveillance of Norway's leaders detected; Aftenposten, 16 Dec 2014. http://www.aftenposten.no/nyheter/iriks/Secret-surveillance-of-Norways-leaders-detected-7825278.html
3. Security Research Labs. https://opensource.srlabs.de/
4. Priority One Security. http://www.p1sec.com/corp/
5. SBA Research. https://www.sba-research.org/
6. Openbts.org. http://openbts.org/
7. The Osmocom (Open Source Mobile Communication) project. http://openbsc.osmocom.org/trac/wiki/OsmocomOverview
8. Android IMSI-Catcher Detector (#AIMSICD). https://secupwn.github.io/Android-IMSI-Catcher-Detector/
9. Dabrowski, A., Pianta, N., Klepp, T., Mulazzani, M., Weippl, E.R.: IMSI-catch me if you can: IMSI-catcher-catchers. In: Annual Computer Security Applications Conference (ACSAC). ACM 978-1-4503-3005-3/14/12 (2014)
10. Daehyun Strobel: IMSI Catcher, Chair for Communication Security, Ruhr-Universität Bochum, 13 July 2007
11. The Osmocom (Open Source Mobile Communication) project – OsmocomBB. https://osmocom.org/projects/baseband
12. Mitchell, T.M.: Machine Learning. Mcgraw-Hill Companies Inc., New York (1997). ISBN ISBN-0-47-042807-7
13. Yegneswaran, V., Giffin, J.T., Barford, P., Jha, S.: An architecture for generating semantics-aware signatures. In: Proceedings of the 14th USENIX Security Symposium, pp. 97–112 (2005)
14. van Do, T., Nguyen, H.T., Momchil, N., Do, V.T.: Detecting IMSI-catcher using soft computing. In: Berry, M.W., Mohamed, A.H., Yap, B.W. (eds.) SCDS 2015. CCIS, vol. 545, pp. 129–140. Springer, Heidelberg (2015). ISSN 1865-0929 ISSN 1865-0937 (electronic)
15. GSMK CRYPTOPHONE. http://www.cryptophone.de/en/
16. Security Research Labs: SnoopSnitch app. https://play.google.com/store/apps/developer?id=Security+Research+Labs&hl=no
17. Vert, R., Vert, J.-P.: Consistency and convergence rates of one-class SVMs and related algorithms. JMLR **7**, 817–854 (2006)
18. Aftenposten data set. http://www.aftenposten.no/meninger/kommentarer/Derfor-publiserer-Aftenposten-hele-datagrunnlaget-for-mobilspionasje-sakene-7849555.html
19. Anomaly Detection algorithm from Twitter. https://github.com/twitter/AnomalyDetection