HyRiM

Hybrid Risk Management for Utility Networks

Collaborative Project

**Deliverable 4.3**

How to Enhance Perimeter Security Using New Surveillance Technologies

Due date of deliverable: [31.01.2017]
Actual submission date: [31.01.2017]

Start date of project: April 1, 2014
Duration: 36 months

Organisation name of lead contractor for this deliverable
University of Passau

<table>
<thead>
<tr>
<th>Dissemination Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PU</td>
<td>Public</td>
</tr>
<tr>
<td>PP</td>
<td>Restricted to other programme participants (including the Commission Services)</td>
</tr>
<tr>
<td>RE</td>
<td>Restricted to a group specified by the consortium (including the Commission Services)</td>
</tr>
<tr>
<td>CO</td>
<td>Confidential, only for members of the consortium (including the Commission Services)</td>
</tr>
</tbody>
</table>
**Deliverable 4.3 HyRiM How to Enhance Perimeter Security Using New Surveillance Technologies**

**HISTORY**

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Reason</th>
<th>Reviewed by</th>
</tr>
</thead>
<tbody>
<tr>
<td>00.01</td>
<td>05-07-2016</td>
<td>First draft version (Initial structure)</td>
<td>Ali Alshawish - UNIPASSAU</td>
</tr>
<tr>
<td>00.02</td>
<td>03-08-2016</td>
<td>Inclusion of partners’ contributions</td>
<td>Ali Alshawish – UNIPASSAU</td>
</tr>
<tr>
<td>00.03</td>
<td>13.01.2016</td>
<td>Integration of final contributions</td>
<td>Zhiyuan Sui – UNIPASSAU</td>
</tr>
<tr>
<td>00.04</td>
<td>20-01-2017</td>
<td>Internal review and improvements</td>
<td>Antonios Gouglidis – ULANC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Syed Asad Ali Naqvi – ULANC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Paolo Giacalone – AKHELA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Massimiliano Turco – AKHELA</td>
</tr>
<tr>
<td>01.00</td>
<td>30.01.2017</td>
<td>Final version</td>
<td>Ali Alshawish – UNIPASSAU</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Zhiyuan Sui – UNIPASSAU</td>
</tr>
</tbody>
</table>

**AUTHORS LIST**

<table>
<thead>
<tr>
<th>Organization</th>
<th>Name</th>
<th>Email and phone number</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNIPASSAU</td>
<td>Ali Alshawish (<a href="mailto:ali.alshawish@uni-passau.de">ali.alshawish@uni-passau.de</a>; phone number: +49 851 509 3054)</td>
<td></td>
</tr>
<tr>
<td>UNIPASSAU</td>
<td>Amine Abid (<a href="mailto:amine.abid@uni-passau.de">amine.abid@uni-passau.de</a>; phone number: +49 851 509 3054)</td>
<td></td>
</tr>
<tr>
<td>UNIPASSAU</td>
<td>Zhiyuan Sui (<a href="mailto:zhiyuan.sui@uni-passau.de">zhiyuan.sui@uni-passau.de</a>; phone number: +49 0851 509 3055)</td>
<td></td>
</tr>
<tr>
<td>UNIPASSAU</td>
<td>Xiaobing He (<a href="mailto:xiaobing.he@uni-passau.de">xiaobing.he@uni-passau.de</a>; phone number: +49 0851 509 3055)</td>
<td></td>
</tr>
<tr>
<td>UNIPASSAU</td>
<td>Hermann de Meer (<a href="mailto:Hermann.deMeer@uni-passau.de">Hermann.deMeer@uni-passau.de</a>; phone number: +49 0851 509 3051)</td>
<td></td>
</tr>
<tr>
<td>AIT</td>
<td>Bernhard Strobl (<a href="mailto:Bernhard.Strobl@ait.ac.at">Bernhard.Strobl@ait.ac.at</a>)</td>
<td></td>
</tr>
<tr>
<td>AIT</td>
<td>Andreas Opitz (<a href="mailto:Andreas.Opitz@ait.ac.at">Andreas.Opitz@ait.ac.at</a>)</td>
<td></td>
</tr>
<tr>
<td>AIT</td>
<td>Stefan Rass (<a href="mailto:stefan.rass@syssec.at">stefan.rass@syssec.at</a>)</td>
<td></td>
</tr>
<tr>
<td>ETRA</td>
<td>Alberto Zambrano Galbis (<a href="mailto:azambrano.etraid@grupoetra.com">azambrano.etraid@grupoetra.com</a>)</td>
<td></td>
</tr>
</tbody>
</table>
Table of Contents

EXECUTIVE SUMMARY ........................................................................................................... 8
1 ABBREVIATIONS .................................................................................................................. 9
2 INTRODUCTION ................................................................................................................... 9
3 EXTENDED PERIMETER OF UTILITY NETWORKS ............................................................. 11
  3.1 INTRODUCTION ............................................................................................................... 11
  3.2 DE-PERIMETERISATION VERSUS RE-PERIMETERISATION .............................................. 12
  3.3 DEFINITION OF PERIMETER ......................................................................................... 13
  3.4 UNDERSTANDING THE STRUCTURE OF A SECURITY PERIMETER ................................. 14
  3.5 UNDERSTANDING THE NATURE AND THE FUNCTION OF A SECURITY PERIMETER ......... 15
  3.6 NON-EXTENDED PERIMETER VS. EXTENDED PERIMETER ........................................... 16
  3.7 IDENTIFYING EXTENDED PERIMETER COMPONENTS IN A CRITICAL INFRASTRUCTURE .................................................................................................................. 17
    3.7.1 Unattended infrastructure ....................................................................................... 17
    3.7.2 Trends and technology populism ............................................................................. 17
    3.7.3 Outsourcing ............................................................................................................. 18
    3.7.4 Human factor .......................................................................................................... 19
    3.7.5 Summary .................................................................................................................. 19
4 SURVEILLANCE TECHNOLOGIES ...................................................................................... 21
  4.1 OVERVIEW OF SURVEILLANCE SYSTEMS .................................................................. 21
  4.2 MOBILE ID CHECK ......................................................................................................... 21
    4.2.1 Motivation ............................................................................................................... 21
    4.2.2 Secure Mobile Identification Check ......................................................................... 21
    4.2.3 Used Secure Elements (TPM and Secure Boot) ...................................................... 22
    4.1.4 Demo system elements .......................................................................................... 22
    4.1.5 Demo system workflow ......................................................................................... 22
  4.3 3D SURVEILLANCE CAMERA ...................................................................................... 24
    4.3.1 Motivation ............................................................................................................... 24
    4.3.2 Passive stereo depth sensing .................................................................................. 24
    4.3.3 Algorithmic framework .......................................................................................... 25
    4.3.4 Evaluation plan and deployment ............................................................................. 28
  4.4 ETRAFIC ......................................................................................................................... 28
    4.4.1 Traffic Management-related data ............................................................................ 29
    4.4.2 Incident detection .................................................................................................... 31
    4.4.3 Operation mode ........................................................................................................ 33
    4.4.4 Application in surveillance security ......................................................................... 34
  4.5 ALERT MONITOR ............................................................................................................ 34
    4.5.1 Alert operator interface ........................................................................................... 34
    4.5.2 Integration of sensors ............................................................................................. 35
    4.5.3 Application in surveillance security ......................................................................... 36
5 CHALLENGES OF SURVEILLANCE TECHNOLOGIES .................................................... 36
  5.1 SECURITY AND PRIVACY THREATS IN SURVEILLANCE SYSTEMS ........................... 36
  5.2 LEGAL AND REGULATION ASPECTS ............................................................................ 38
  5.3 TECHNICAL SOLUTIONS ............................................................................................... 40
  5.4 OPEN QUESTIONS .......................................................................................................... 41
6 HOW TO ENHANCE SECURITY USING SURVEILLANCE TECHNOLOGIES ........................................... 42

6.1 MOTIVATION (LIMITATIONS OF TRADITIONAL SURVEILLANCE SYSTEMS) ................................. 42

6.2 GAME-THEORETIC APPROACH FOR SURVEILLANCE ENHANCEMENT ........................................ 42
  6.2.1 Standard Model ....................................................................................................................... 42
  6.2.2 Game Theory using Uncertainty ............................................................................................. 43

6.3 SECURITY ENHANCEMENT USING VIDEO SURVEILLANCE SYSTEMS ........................................... 46
  6.3.1 Introduction ............................................................................................................................ 46
  6.3.2 Pattern Retrieval from surveilled Image Data ....................................................................... 46
  6.3.3 Pattern Clustering and Risk Level Assignment ...................................................................... 47
  6.3.4 Pattern Prediction based on Particle filtering and SIFT Features ........................................ 49
  6.3.5 Application of Game-theoretical approach (HRMs) to support Video Surveillance ............. 50

6.4 OPTIMAL SURVEILLANCE SCHEDULE BASED ON MOBILE ID CHECK TECHNOLOGY ................... 52
  6.4.1 Simulation ............................................................................................................................. 52
  6.4.2 Scenario ................................................................................................................................ 58

6.5 A PRIVACY BREACH EVALUATION MODEL ON MOBILE ID CHECK TECHNOLOGY ....................... 66
  6.5.1 Overview of Location Privacy .................................................................................................. 67
  6.5.2 Previous Evaluation Approaches ............................................................................................ 68
  6.5.3 Assumptions ........................................................................................................................... 68
  6.5.4 Notations ................................................................................................................................ 68
  6.5.5 Attackers ................................................................................................................................. 69
  6.5.6 Evaluation Model .................................................................................................................... 69

7 CONCLUSIONS .................................................................................................................................. 78

8 REFERENCES ....................................................................................................................................... 79

ANNEX I: QUESTIONNAIRE FOR ELICITING EMPLOYEES’ PERCEPTION OF PRIVACY ....................... 84

ANNEX II: PAYOFF STRUCTURES FOR GOALS DAMAGE, DETECT RATE, AND PRIVACY ...................... 85
Index of Figures

Figure 1 Overview of various components of an extended and non-extended perimeter ........................................ 20
Figure 2 System elements: Server with access permissions of persons, Secure ID check device ............................ 22
Figure 3 Enrollment of person ....................................................................................................................... 23
Figure 4 Verification of person ....................................................................................................................... 23
Figure 5 Stereo sensor heads of 40 cm baseline (left) and 20 cm baseline (right) ............................................ 24
Figure 6 Computation scheme of the employed sparse spatio-temporal features D = (D_1, . . . , D_m), forming a local structural representation. Red cuboids illustrate some sample sparse structural units which can be used to represent the local spatio-temporal gradient structure .................................................................................................................. 26
Figure 7 Illustration showing the testing stage of a previously unseen input signal. A set of (at training time) learned dictionary entry combinations are employed to test the reconstruction quality of the input signal. If the best reconstruction from K combinations delivers a low error, it indicated that the input signal is well represented and it can be well approximated by the learned sparse code combinations. Otherwise, high reconstruction error indicates that similar signals have not been encoded in the testing combinations, and signals an unusual input signal ........................................................................................................ 27
Figure 8 Left: table listing the staged unusual events for a selected indoor scenario. Right: a screenshot of the staged queue (normal behavior) and a loitering scenario (red bounding box annotation), and its equivalent top-view representation (inset B) ........................................................................................................ 28
Figure 9 Video recorder .................................................................................................................................. 29
Figure 10 Example of video image .................................................................................................................. 29
Figure 11 Virtual loop ..................................................................................................................................... 30
Figure 12 Queues ............................................................................................................................................. 30
Figure 13 Lane occupation ............................................................................................................................... 31
Figure 14 Excessive queue ............................................................................................................................... 31
Figure 15 Stopped Vehicle ............................................................................................................................... 32
Figure 16 Vehicle with emergency lights ........................................................................................................ 32
Figure 17 Excess/Defect of traffic intensity ...................................................................................................... 33
Figure 18 Openation model ............................................................................................................................. 33
Figure 19 Operator interface ............................................................................................................................ 34
Figure 20 Amount of alerts triggered during the last 24 hours ........................................................................ 35
Figure 21 Details per sensor ........................................................................................................................... 35
Figure 22 Threats Model on Security and Privacy in Surveillance Systems ..................................................... 37
Figure 23 Procedures of the BOF SIFT algorithm ............................................................................................ 47
Figure 24 Occurrences of clustered patterns .................................................................................................. 49
Figure 25 Example patterns from the first cluster (where SIFT keypoints are shown) ...................................... 49
Figure 26 Probability-distributed risk levels .................................................................................................. 50
Figure 27 Loss distribution of the sample surveillance system ....................................................................... 51
Figure 28 Site map sample ............................................................................................................................... 53
Figure 29 Modeling of paths ............................................................................................................................ 53
Figure 30 A built physical environment sample .............................................................................................. 54
Figure 31 Actors .............................................................................................................................................. 55
Figure 32 current strategy to select an area for Check ...................................................................................... 56
Figure 33 screenshot of the simulated area ....................................................................................................... 58
Figure 34 the comfort breach curve ................................................................................................................. 59
Figure 35 Optimal Defense Strategy ............................................................................................................... 64
Figure 36 Worst-case Attack Strategy ............................................................................................................. 64
Figure 37 Assurance Plots of Damage ........................................................................................................... 65
Figure 38 Assurance Plots of Detect Rate ........................................................................................................ 65
Deliverable 4.3 HyRiM How to Enhance Perimeter Security Using New Surveillance Technologies

Figure 39 Assurance Rate of Privacy ................................................................. 66
Figure 40 Communication model of privacy breach with external attackers ........................................ 69
Figure 41 Communication model of privacy breach with internal attackers ........................................ 73
Figure 42 Probability of an Event ................................................................. 73
Figure 43 Two State Markov Chain Representing Subject Movement From/To a given area ............ 74
Figure 44 Transient Regime ........................................................................ 75
Figure 45 Convergence time (Δt) as a function of λ ........................................ 76
Figure 46 $P(t)$ as a function of time for different values of $\lambda$ ........................................ 76
Figure 47 Loss distribution (Payoff structure) for goal “Damage” ......................... 85
Figure 48 Loss distribution (Payoff structure) for goal “Detection Rate” .................. 86
Figure 49 Loss distribution (Payoff structure) for goal “Privacy” ........................................ 87
Deliverable 4.3 HyRiM How to Enhance Perimeter Security Using New Surveillance Technologies

Index of Tables

Table 1 Transitions in the threats model......................................................... 37
Table 2 Payoff matrix of the sample surveillance system ................................ 51
Table 3 Security levels .................................................................................. 58
Table 4 simulation set up (General parameters) ............................................... 60
Table 5 Defender strategies ........................................................................... 60
Table 6 Attacker strategies ............................................................................ 60
Table 7 Considered strategies ....................................................................... 61
Table 8 Categories ......................................................................................... 62
Table 9 D-NGXFXT vs. A-NIXTR ................................................................. 62
Table 10 D-NGXFTHSLF vs. A-NIXTR .......................................................... 63
Table 11 Location Privacy Unfriendly Technologies ....................................... 67
EXECUTIVE SUMMARY

Surveillance plays an important role in the protection of critical infrastructures. However, current surveillance systems are rigid and inefficient because the deploying is a long, complex process. The HyRiM deliverable D4.3 studies the perimeter security, introduces new state-of-the-art surveillance technologies for perimeter security and explain in details how those technologies can be used to enhance existing surveillance systems. The study of perimeter security examines various definitions of the perimeter. Following, important elements of an extended perimeter are identified. Due to the importance of surveillance systems for ensuring the security of a perimeter, some new surveillance technologies are introduced (i.e., 3D surveillance and Mobile ID check). Firstly, 3D surveillance is a passive stereo depth vision sensor, which provides surveillance images in three dimensions. As opposed to two dimensions surveillance data – a technique that is unsuitable for depth decays with increasing distance -- 3D techniques can take care of the occlusion problem and can distinguish multiple objects from the scene quite accurately. Secondly, Mobile ID check is a portal and wireless on-demand surveillance device. The approach is advanced in technology beyond state of the art in biometrics data and badge scanning. It brings a benefit for all general identity checks with legal badges which can be read by a dedicated handheld device. By using it, a physical access control system can be enhanced. Security guards move around the monitored area and check subjects’ identities according to predefined strategies. Working with Alert Monitoring, which is an interface to automatically trigger alert, the new surveillance technologies can provide clear real-time information of the events happening in the different parts of a system. Subsequently, we analyse possible vulnerabilities of these technologies and enumerate some available technical solutions. However, these solutions cannot eliminate all security and privacy risks in the system. It is important to give recommendation referring to current EU data protection regulations considering those risks. Based on the recommendations, D4.3 proposes two models on the security and privacy on surveillance systems. The suspicious behavior prediction model is based on data retrieval and particle filtering. To enhance security using video-based surveillance systems, Scale Invariant Feature Transform (SIFT) is applied to retrieve risk relevant patterns in images. Furthermore, an improved version of K-means clustering method is proposed to cluster retrieved risk-related patterns according to a risk level assignment method of the security operator of the video surveillance system. Afterwards, the game-theoretical framework, developed in WP1, are applied on those patterns to give recommendations on the optimal configuration of the video surveillance system. The privacy evaluation model, based on entropy theory, outputs possible privacy impacts of new on-demand surveillance techniques. The privacy evaluation model involves asynchronous checks. A surveilled subject’s original location may affect the probability of events. Therefore, a probability prediction model is designed based on Markov chain model in order to predict probabilities of events affected by adversaries’ experience. Then, based on entropy, subjects’ location uncertainty is measured. Finally, in order to show the functionalities of game theory in Mobile ID check, a simulation model is constructed to provide patterns to find optimal surveillance strategies. Based on the simulation model, a scenario is designed to demonstrate that with different metrics values from the simulation model, the game can decide on the optimal strategies to instruct security guards. Furthermore, the respective optimal loss distributions attained under any behaviour of the attackers are given.
1 ABBREVIATIONS

<table>
<thead>
<tr>
<th>Term</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>Access Control</td>
</tr>
<tr>
<td>ABC</td>
<td>Automated Border Control</td>
</tr>
<tr>
<td>ACE</td>
<td>Akhela Code of Ethics</td>
</tr>
<tr>
<td>AOR</td>
<td>Areas of Responsibility</td>
</tr>
<tr>
<td>B2B</td>
<td>Business to Business</td>
</tr>
<tr>
<td>B2C</td>
<td>Business to Customer</td>
</tr>
<tr>
<td>BDSG</td>
<td>Bundesdatenschutzgesetz (Federal Data Protection Act)</td>
</tr>
<tr>
<td>BoF</td>
<td>Bag of Features</td>
</tr>
<tr>
<td>CI</td>
<td>Critical Infrastructure</td>
</tr>
<tr>
<td>CIP</td>
<td>Critical Infrastructure Protection</td>
</tr>
<tr>
<td>DFT</td>
<td>Discrete Fourier Transforms</td>
</tr>
<tr>
<td>DoG</td>
<td>Difference of Gaussian</td>
</tr>
<tr>
<td>DPA</td>
<td>Data Protection Act</td>
</tr>
<tr>
<td>DSG</td>
<td>Datenschutzgesetz (Federal Act Concerning the Protection of Personal Data)</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>GDPR</td>
<td>General Data Protection Regulation</td>
</tr>
<tr>
<td>HRMs</td>
<td>Hybrid Risk Metrics (Game-theoretical Framework developed in HyRiM)</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hyper Text Transfer Protocol</td>
</tr>
<tr>
<td>ICT</td>
<td>Information and Communication Technology</td>
</tr>
<tr>
<td>IoT</td>
<td>Internet of Things</td>
</tr>
<tr>
<td>M2M</td>
<td>Machine to Machine</td>
</tr>
<tr>
<td>PKI</td>
<td>Public Key Infrastructure</td>
</tr>
<tr>
<td>POC</td>
<td>Phase-Only Correlation</td>
</tr>
<tr>
<td>PS</td>
<td>Perimeter Security</td>
</tr>
<tr>
<td>SIFT</td>
<td>Scale Invariant Feature Transform</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>TPM</td>
<td>Trusted Platform Modular</td>
</tr>
<tr>
<td>WP1</td>
<td>HyRiM’s Work Package 1 (Hybrid Risk Metrics and Methodology for Risk Assessment)</td>
</tr>
<tr>
<td>WP4</td>
<td>HyRiM’s Work Package 4 (Perimeter Protection Enhancements)</td>
</tr>
</tbody>
</table>

2 INTRODUCTION

Critical infrastructure protection (CIP) is increasingly becoming a major concern in governments and industries. Besides the increasing rates of cyber-crime, recent terrorist attacks bring CIP into a severer environment. Perimeters are usually treated as the first line of defense for CIP. Surveillance systems have been used for perimeter security, such as access control and malicious behavior detection. Traditional perimeter security (PS) solutions typically monitor these outer boundary structures and lines, thus ignoring threats from the inside [1]. Due to the inflexibility and fixed installation of these systems, their deterrent effect will be considerably less [2]. The monitoring system predictability is very high and the attacker can observe and adapt to the current surveillance strategy. Therefore, it is important to maintain situational awareness within the industry complex so that the potential intruders can still be detected. The definition of perimeter security should be extended from the boundary of the assets to the outside, e.g., human resources and the outsourcing data. Therefore, novel surveillance devices should be studied to cope with this dynamic nature and to achieve adequate level of situational awareness in such large-scale areas taking account of limited available resources (e.g., security guards and badge check devices). The development of the Information and Communication Technology (ICT) is a promising chance to handle that challenge.
Additionally, more and more portable devices are used today, which enable surveillance strategies to be risk-based and on-demand.

The application of those novel technologies, however, needs quantitative metrics to inform devices operation in an optimal way. In this deliverable, we mainly consider using new surveillance devices in order to enhance current surveillance systems, such as Mobile ID check and 3D surveillance. 3D surveillance is able to deliver dense 3D data in real-time and cope with a wide range of ambient light conditions. With the help of 3D sensing, the operators of the system can virtually reconstruct a 3D scene to detect and tracks unusual behaviors and activities within the context of the targeted surveillance functionalities in order to capture situations that trigger alerts to human operators. Combined with human-machine interaction technologies, the system can highlight the target and provide visual options to employee. In particular, the potential extensions of the surveillance capabilities by utilizing Mobile ID check as additional sensors for the surveillance infrastructure. Mobile ID check is a kind of fast and secure identification device. It is one of the keys to control physical accesses. With it, security guards can dynamically check subjects’ identification badge. In order to improve the security of the device, trusted platform module (TPM) is employed. The encryption key or hash value is stored inside the memory of the device in a firmware. Besides these two devices, D4.3 introduces ETRAFIC, which is a computer vision system capable of analyzing video streams. The system consists of a set of IP cameras, a video recorder and a monitoring system. In order to integrate different surveillance systems, D4.3 also introduce alert monitor. The alert monitor provides interfaces for real-time information of events happening. This would benefit system operators to take required actions as fast as possible.

Novel surveillance systems offer more flexible surveillance approaches. The system operator can set various surveillance strategies. However, the new surveillance devices also introduce more vulnerabilities. Security and privacy are two challenging tasks in new surveillance systems. D4.3 utilizes an attack tree model to analyze possible security and privacy threats in surveillance systems. It covers data collection, processing and storage to minimize privacy outages in on-demand surveillance systems. Therefore, it is important to achieve an insight on the legal compliance and give legal recommendations. In this vein, it surveys some available security and privacy technologies. However, those technologies cannot solve all security challenges. D4.3 enumerates some open challenges related to new surveillance technologies. Those challenges contain threat detection and privacy evaluation. Surveillance devices provide an amount of data to employees. It is impracticable to visualize simultaneously the behavior of all the observed subjects, in order to quickly and correctly detect danger situations. If only abnormal situations are automatically screened and graded, an employee can take corresponding measures in real-time. It would save a lot of time and labor. While subjects usually appreciate the sense of increased security brought by on-demand surveillance, they often fear the loss of privacy that comes along. During surveillance, private information of subjects under surveillance is leaked, including their locations, behaviors and habits [3]. For example, since security guards are able to track the position and movements of monitored employees, they could determine that some employees have not been working as expected, and thus take disciplinary actions. This can be done through analyzing the individual data of the respective employees.

In order to provide patterns that lead to risks that can be quantitatively predicted based on the new surveillance systems, in D4.3 we propose suspicious behaviors prediction and privacy breach evaluation models. In the proposed risk prediction model, firstly, Scale Invariant Feature Transform (SIFT) descriptor is utilized to describe the local patch in images which are obtained from video surveillance systems. Then, an improved version of K-means is introduced to cluster SIFT descriptors and risk levels are assigned to those clusters. Finally, an experimental application is given to demonstrate that the game-theoretical framework (HRMs) can take the results of this model as input to optimize video surveillance configurations. In the context of WP4, an experimental simulation has been developed to evaluate the potential damage, privacy leakage and detection rate in different scenarios involving Mobile ID Check technology. According to the
obtained results, the system operator can quantitatively measure each surveillance strategy and find out the optimal strategy to trade off those security, privacy and accuracy effects. The simulation model consists of three phases. In the first phase, given a strategy, target areas are selected and security guards are sent to those areas. After that, in the second phase, the security guards move the selected areas and check subjects. In the final phase, security guards move back to their headquarters waiting for a new assigned mission. Two privacy evaluation models are proposed in order to provide theory for privacy estimation of surveillance strategies. The basic model is based on Shannon entropy against external attackers. The second model considers persistent attackers. In order to estimate the dependence among asynchronous checks, we set a Markov chain based model. The probability of an event is decreasing as time goes by in the new model.

The remainder of this deliverables is organized as follows: Section 3 defines and extends the definition of security perimeter in utility networks. Section 4 introduces studied surveillance technologies. Section 5 summarizes challenges of surveillance technologies, from both technical and legal aspect. In Section 6, we elaborate on possible security enhancements using video surveillance systems and Mobile ID check technologies. Section 7 concludes this deliverable.

3 EXTENDED PERIMETER OF UTILITY NETWORKS

3.1 Introduction

The vast majority of critical infrastructures (CI), especially utility networks such as power, transportation, water and gas networks, have been working (operational) for several decades. They represent the main pillars for national economy and prosperity, as they provide essential services and fundamental networks upon which we all are extensively dependent and tightly linked. In fact, advances of some infrastructures, in particular telecommunication and transportation, have a significant impact on other sectors by providing new opportunities and mechanisms for improving business efficiency, coordinating large-scale delivery operations and managing complex supply chains [4]. Given the high reliance upon such infrastructure, any outage, inadequate service supply, or even temporal disturbances can adversely impact the quality of life as well as public safety and security and hence nation’s progressiveness and competitiveness. Disruptions of CI are most likely associated with high cost due to the potential delay in delivering services and cost of damage recovering. Moreover, the damage is not always limited to the affected area but can easily propagate into other sectors and areas. Therefore, securing and protecting CI has been addressed as a national priority in many countries. Additionally, the resilience of CI attracts more global attention and considerable interest in both the in industrial and scientific worlds.

Intuitively, protection implies the state of keeping the valuable assets, which an organization owns, manages, or controls, from being damaged, stolen, or lost. Hence, those assets that are of concern to an organization are well identified or at least obviously identifiable. Whether done explicitly or not, defining the perimeter of an organization can considerably simplify the process of identifying these valuable resources not only for effective protection strategies but also for preparing proper coordination and management plans and policies. Furthermore, having clear boundaries serves to identify the organization’s scope of responsibilities including the wide range of activities and duties to absorb potential disturbances and the ability to respond and recover from failures and disasters. Given the fact that CI are increasingly becoming interconnected and interdependent, the role of the organization’s perimeter is becoming more vital and important to recognize the different linkages and interconnections and hence to avoid being damaged by other interconnected systems. By addressing this issue, organizations’ resilience and robustness can be enhanced and the confidence in infrastructures can be boosted, as well.
CI and their respective assets and facilities usually tend to feature a significant size, spreading over large areas and across long distances connecting regions, which are geographically far apart. These infrastructures can also cross regional and national boundaries passing through different environments and different environmental circumstances. These environments can be characterized by different national, global, personal, organizational, business and operational aspects. Consequently, the large geographic span of such systems can be solely responsible for making the process of defining, visualizing or even protecting the respective perimeter an extremely challenging task.

There is another perspective that has to be considered while identifying an organization’s perimeter; namely the virtual limits of organizations such as implemented cyber-infrastructures. Critical infrastructures exploit ongoing advances in information and communication technologies (ICT) for supporting control and automation of their processes and enhancing the ability to adapt to changes rapidly and subsequently to achieve resilient operation performance. The cyber and computational elements can include control systems of physical infrastructure together with business and corporate network infrastructure. In fact, this issue is of vital important since these systems can extend beyond its physical borders to include other entities such as vendors, business partners, service providers or even costumers. As a result, various elements or even systems, which were previously isolated from each other by clear and well-defined boundaries, might be spontaneously and seamlessly integrated into one system crossing the boundaries marked by their traditional individual perimeters.

Consequently, boundaries between systems are increasingly becoming fuzzy or even vague. We emphasize on the difference between fuzziness and vagueness, based on Zadeh’s claim in [5], who argues that there is logically a significant difference between fuzziness and vagueness. The former implies unsharpness of boundaries, which means the borderline between two entities still exists but it is inexact or imprecise (i.e. the boundaries can be identified even within a specific range implying the possibility to be (partial) controllable). Conversely, vagueness connotes lack of specificity, which means inability of making decision due to lack of sufficient information (i.e. boundaries between two entities are not identifiable, and hence incontrollable anymore). Furthermore, development and deployment of perimeter security measures are becoming more complex and expensive than ever before, while penetration means are getting cheaper and publicly available [6].

3.2 De-perimeterisation versus Re-perimeterisation

Nowadays, agility and flexibility are key characteristics of organizations, which evolve in a complex and uncertain business environment. As a result, the process of blurring or breaking down boundaries threatens such contemporary organizations more and more. This process was also referred to as “De-perimeterisation”, originally coined by a former chief security researcher at UK’s Royal Mail Group - Jon Mescham [7]. Afterwards, this term was adopted and promoted by the Jericho Forum, an international working group hosted by Open Group, established to deal with the challenges associated with surviving in a network without boundaries. Therefore, the Jericho Forum refers to de-perimeterisation as a concept or strategy that uses a mixture of inherently-secure protocols and components to protect organizations’ data and systems rather than the reliance on the security boundary to the Internet. Simultaneously, the term of de-perimeterisation is used to describe the process of a gradual dissolving of an organization’s security perimeter focusing only on the cyber world and data protection [8].

Due to the current trend of networked world, it is likely that the organizations do not have their own IT infrastructure or even do not have any control over it. Nowadays, complex systems are built on top of other systems, and probably communicate with other ones [9]. Therefore, unknown and obscured connections and dependencies become more likely. This fact becomes particularly inevitable after emerging new business paradigms such as Business-to-Business (B2B), Business-to-Customer (B2C) and Machine-to-Machine (M2M).
In this case, having a perimeter as a protective measure would significantly impact the level of connectivity and hence strongly impede the envisaged business growth and collaboration. In other words, the security perimeter would act more in a blocking manner rather than in a facilitating or enabling one with respect to business objectives.

The prevalent trend towards more connectivity and emerging technologies, such as cloud computing, virtualization, Internet-of-Things (IoT) and mobile internet, constitute the essential factors behind vanishing boundaries. The various benefits associated with the increases in connectivity; such as cost saving, flexibility, enabling E-business, enhanced efficiency, improved management and coordination, among others, are vital forces in stimulating adoption and acceptance of blurring boundaries. Furthermore, there are several technical and organizational mechanisms that have detrimentally affected the organizations’ security perimeter such as increased network capacity, increased assets’ mobility, transferring encrypted data, business and employee dynamics, service oriented application, individual empowerment, among others [10]. However, there are several other forces opposing this trend, including need for accountability, privacy, reliability, safety and security that places an inevitable demand due to the rapid increase of security incidents in critical infrastructures. According to Cleef et al. [9], these forces are pushing on the opposite end against de-perimeterisation, leading to re-perimeterisation. In other words, valuable assets have to be surrounded by a protective perimeter.

Nevertheless, defining the assets to be protected is a challenging task. Thus, security measures are increasingly placed at lower and lower level ending up with defining the perimeter solely around the data rather than around the entire infrastructure [9]. It is extremely difficult, if not impossible, to evaluate security only at the data level since data represents the end result of various processes at different higher levels; such as technical, individual and organizational [10]. Moreover, blurring boundaries have a negative impact on the compliance with (and enforcement of) regulation, as well. For this very reason and because of the factors outlined above, there is an inevitable and urgent necessity for a proper definition of a perimeter as well as the classification of assets, interactions and responsibilities.

3.3 Definition of Perimeter

It is worth looking at some definitions of perimeter in an attempt to understand its value to existing systems and networks. The Cambridge dictionary, for example, defines the term perimeter as

“The outer edge of an area of land or the border around it” [11]

By contrast, the American Heritage Dictionary defines it as

“A defended boundary protecting a military position” [12]

Obviously, both definitions are subject to interpretations varying according to the considered context. Nevertheless, we can at least figure out the basic functionality of a perimeter from these definitions. Intuitively, there are two main functions that are manifestly stated by both definitions and will immediately come into mind when you hear the term “perimeter”; demarcation and protection, of course after excluding the mathematical intuition. Demarcation refers to the process of setting or marking boundaries or limits [12]. In turn, protection is associated with the preservation from harm, destruction, and loss as well as unwanted activities. By combining both aspects we come to a conclusion that the perimeter is the outer boundary that protects the inside (holdings) from the outside (danger). This conclusion comes in line with Cleeff’s definition of security perimeter [10]

“A security perimeter is a technical solution to protect assets from negative influences originating in its environment.”
According to this definition, sources of threats and negative influences to an arbitrary asset are located mainly in its surrounding environment, which is referred to as “outside”. Since the perimeter is a borderline, at which the inside ends and the outside begins, it is also the first entity that any external entity from the outside has to come into contact with prior to infiltration into inside. Depending on this discussion we propose the following definition of the term perimeter:

**The perimeter is any assembly or construction that physically or virtually surround a facility, its various assets, (or even a simple predefined area); through it the contact (communication paths) between the internal world, i.e. the entity to be protected, and the external world or environment, will be established and facilitated.**

Organizations, especially those that are classified as complex and coupled systems such as utility systems, usually feature a multilayer architecture including physical infrastructure layer, control layer, business layer, among others. Hence, each individual layer has its own boundaries not only with the external world but also within the same organization. In turn, this results in creating several boundary and perimeter components of different natures, such as:

- Organisational perimeter
- Geographical or physical perimeter
- Business perimeter
- Cyber perimeter
- Data perimeter: *while the cyber perimeter can define the boundary between private-owned network devices and the public ones, some organisations can, for some legal reasons, transcend the cyber boundary to match the national ones.*
- Legal perimeter
- Individual perimeter
- Security perimeter
- etc.

Consequently, the security perimeter of an organization, which is responsible for provisioning protection at different layers, should be hybrid and multi-dimensional to be able to reduce physical and cyber risks as well as business risks.

### 3.4 Understanding the structure of a security perimeter

Based on the definition we presented above, the security perimeter can be presented as a component or a (sub)system that is responsible, firstly, for maintaining the required isolation and separation between the internal private zone and the external public world (i.e. the perimeter encompasses the solid defending wall surrounding the inside). Secondly, it is responsible for provisioning the appropriate paths of communication from and to the internal zone (i.e. the perimeter embraces the doors allowing access to the inside). Thus, the core components of a security perimeter would be:

1. **A wall**: it surrounds the entity to be protected tearing apart the area (space) into two areas, inside and outside. The inside refers to the internal area behind the wall and towards the target entity. In contrast, the outside refers to the external area, which starts at the wall and spans away from the target. Thereby, the wall maintains inaccessibility of the inside from the outside and vice versa. Therefore, the wall’s function is to **always** prevent access regardless of the circumstances.
2. **A door(s)**: it represents a controlled and monitored pathway for entering and leaving the internal zone. The door plays two functional roles:
   a. In case of unauthorized entities, it integrates to the defending wall to maintain isolation (closed mode).
b. Whereas in case of legitimate entities, it provides the way to circumvent the surrounding solid wall (open mode).

In other words, a door is (should be) simply a wall for those without a proper key. Through doors diverse permissible communication paths from and to the internal zone are provisioned and mediated.

In addition to door and wall components, there is another element called **Hole** (unwanted but most likely not entirely avoidable). A hole refers to a location where the wall fails to fulfil its mission in maintaining inaccessibility. Holes are attractive elements for potential intruders since they are always-open doors with open (unauthorized) access connections.

In a nutshell, the efficiency and effectiveness of any security perimeter, either physical or cyber, mainly depends on:

1. How well the wall surrounds the system of interest? It addresses the full coverage feature of the perimeter.
2. How well the doors block or filter out the malicious entities and traffic? It can be measured in terms of false positives and false negative rates.
3. How many (unpatched) holes are there in the wall as well as in the door, as the door plays the role of a wall in face of illegal entities? This is more challenging to measure since holes refer undoubtedly to hidden pathways and unknown entry points into the internal network. On the contrary, this fact raises attention on the importance of performing routine testing and assessments of the system and its perimeter to identify and uncover holes and potential weaknesses.

### 3.5 Understanding the nature and the function of a security perimeter

Security perimeter is responsible for keeping adversaries and malicious entities away, on the one hand, and preventing or reducing loss, leakage and theft of assets and resources such as sensitive data, on the other hand. Simultaneously, the perimeter plays also an important role in protecting the external world from damages stemming from the internal world. The perimeter as the main provider of contact points defines the components that constitute the organization’s attack surface. These components have to be appropriately controlled and managed in order to achieve the envisaged protection and security. The contact points (paths) are normally provisioned by door-entities that are separated from the protected assets, for example by firewalls, VPN gateways, NAC devices in the cyber world as well as gates for authorized ingress and egress to the respective facility, biometric access control systems and security guards in the physical world. These entities are primarily employed to provide a valuable protection and control capabilities. From a strategic point of view, they are deployed only at borders of the protected site serving as the first line of contact and defence. In spite of all that, some system elements can immediately communicate with external elements without any mediation through the deployed security perimeter; for example control devices with wireless internet connection, resources delivery network infrastructure, mobile devices, human resources, or even data itself. Control devices with data or internet connections can provide paths for data to move and migrate or provide outsider access to the internal network. Delivery and transportation grids of the provisioned resources; such as water and gas pipeline or power transmission and distribution networks, are ultimately difficult to be confined behind a dedicated perimeter infrastructure since they often extend for very long distances. Mobile devices and human resources cannot be consistently confined behind static and predefined perimeter structure, as well. Data can also have an unmediated contact with the external untrusted environment since it might be exported and exchanged with other organizations or individuals for the purposes of support and collaboration. In this way and according to the definition (See Section 3), these

---

1. Virtual Private Network
2. Network Access Control
elements have to be immediately moved to be part of the perimeter since they are directly in contact with the external world entities and have to be properly controlled, monitored and managed. Otherwise, these elements will result in numerous holes in the protective perimeter, rendering it more porous and less effective. It is not difficult to imagine the number of potential holes in an organisation’s perimeter with one thousands employees in a workplace, and everyone is equipped with his tablet or smartphone, which can easily bridge the gap between external and internal zones through the available simultaneous access to corporate network and public network in the same device. In order to avoid any potential confusion and to consider the new dynamic nature of the perimeter due to the ongoing technological development, we would like to go further and divide the perimeter into two subgroups; non-extended and extended perimeter. The former refers to the diverse entities that aim at ensuring security and safety of a system, while the latter oppositely refers to potential attack vectors. Hence, there is a vital need to pay special attention towards extended perimeter components.

3.6 Non-extended perimeter vs. extended perimeter

Traditionally, following the castle (fortress) metaphor for security, the system perimeter has included common defending components such as fences, guards, firewalls, IDS, etc. This type of perimeter has also been designed and implemented with the D5 strategy in mind; i.e. Demarcation, Deter, Detect, Delay and Defend [13]. Demarcation refers to the process of creating virtual and physical boundaries around facility’s assets, such as buildings and data. In the context of physical security, these boundaries should be also visible to avoid innocent boundary crossings and to simplify identifying hostile intentions. The goal of deterrence is to create an unattractive environment for potential adversaries. Security lighting, fences, monitoring points, and surveillance systems are, for example, effective deterrents since they are able to reduce the attacker’s opportunity to commit his attack unobserved. However, deterrent efforts are not enough to keep adversaries out. Therefore, it is of vital importance for the perimeter to be able to detect unwanted activities and to delay potential perpetrators long enough to allow security forces or first responder to intercept and defend by denying access to the internal critical assets and resources. In the context of physical and cyber security, the perimeter can include fences, walls, monitoring points, entrance gates or doors, vehicle barriers, security lighting, landscaping, surveillance systems, alarm systems, gates, firewalls, routers, access control devices, intrusion detection systems, among others. All of these measures have been used to ensure that any contact with the protected assets is authorized by the predefined security perimeter before taking place within the internal zone. Due to the apparent static nature and predictable placement of these mechanisms, it is probably valid to refer to them as the “non-extended perimeter”.

In contrast to non-extended perimeter components, including diverse monitoring and controlling mechanisms provisioning security functions at the predefined system boundaries, there are still other entities, assets and system elements, which belong to the infrastructure to be protected by the non-extended perimeter, but they are directly accessible from the external world without any intermediate mechanisms controlling and managing the communication paths. In other words, the paths of communication are directly provisioned and controlled by the communicating entities themselves. Thereby, these entities most likely have the ability to provide hidden and unobserved communication paths that can bypass and circumvent the traditional non-extended perimeter infrastructure. In such case, these elements have to move directly to the perimeter constituting an extension to the common (non-extended) perimeter since they provision communication paths with the outside. Occasionally, some of these elements can presumably offer or ensure a certain level of security. Nevertheless, the vast majority of them poses certainly serious threats to the system of interest. Dynamic system components, such as mobile devices, USB sticks, outsourced data and human resources, are good examples of the entities that have to be permanently considered as parts of the

---

3 Intrusion Detection System
**extended perimeter** since their movement and their associated communication paths are not always foreseen and comprehensible. The components of this group are almost involved in two roles; operational role as a part of the business and operational process and perimetric role due to the ability to immediately communicate with the external world. Furthermore, the extended perimeter elements have to be to some extent owned, managed, or controlled by the organization, and they directly or indirectly impact its business and operational processes and/or the organization has liability for the damage caused by their errors, failures or malfunctions.

### 3.7 Identifying extended perimeter components in a critical infrastructure

Broadly speaking, an organization’s extended perimeter encompasses all components and means that have potential for circumventing the protective borders as well as bridging the air-gap maintained by the non-extended perimeter. They are ordinarily legitimate components allowed to pass through the perimetric doors (i.e. conventional perimetric security measures) during the day-to-day system operations. Hence, they pose a high risk for security, especially since they can serve as potential carriers of infection for other interconnected entities. As a result, we can basically identify four elements of an extended perimeter.

#### 3.7.1 Unattended infrastructure

This group encompasses all elements which are not explicitly or implicitly behind any non-extended perimeter infrastructure, and they are obviously situated in a potentially hostile environment. For examples:

- **Transportation grids for the provided resources**: Examples of such elements are water and gas pipelines as well as power transmission and distribution networks. Their sheer scale of spread makes it almost impossible to deploy and manage a typical defending perimeter around them. Unattended On-field control stations are also elements of this group.
- **Control system devices**: These devices are increasingly outfitted with networking capabilities to allow remote access and configuration. Thereby, they are directly accessible from outside through modem access, DSL and ISDN lines, wireless access, and/or VPN tunnels that can be easily exploited to bypass firewall because of its encrypted nature.
- **Mobile delivery entities**: fuel delivery vehicles as well as bulk oil delivery tanker trucks and vessels are good example of these elements. For attackers, these entities are relatively easy targets with high potential of disruption to the respective supply chain. Going beyond, they can be also leveraged by adversaries against other potential targets due to their mobile nature.

These elements represent a serious concern to critical infrastructure operators as the most popular targets for attackers. The specific nature of these entities make attacks more dramatic and devastating. For example, the large economic loss and environmental effects associated with an explosion in a major oil or gas pipeline.

#### 3.7.2 Trends and technology populism

Traditionally, an organisation’s IT department is solely responsible for planning and choosing the organisation’s IT infrastructure as well as provisioning employees with necessary hardware and software they need for their daily job. Therefore, training budget is an important and inevitable part of any system planning process, since the employees have to be trained how to use these tools and applications to ensure conveying skills necessary for productivity. Nowadays, however, employees’ productivity increasingly depends on the use of personally owned technologies and systems, in particular for tech-savvy workforce who grew up with
Deliverable 4.3 HyRIM How to Enhance Perimeter Security Using New Surveillance Technologies

technologies. The key force enabling and powering this trend is the usability, ease-of-use of owned systems and IT self-supporting, which outweighs security requirements from the employees’ perspectives. The trend of bringing tools and applications designed to be used at home for personal usage to workplaces for further business purposes is referred to as technology populism [14] Online collaboration, ubiquitous internet, social networking, mobile devices, among other are all means for increased happiness and productivity of employee through enabling IT self-provisioning. On the contrary, technology populism brings security threats and risks alongside its benefits, such as:

- The organisation doesn’t own the whole IT infrastructure used in its business operations and hence they don’t have valid assurance of quality-of-service.
- IT managers are not able to train and to provide support for every tool and application used by the employees. This results in creating a very complex support environment and lack of centralized administration.
- Security managers have neither full control nor full awareness of the current set of application and information resources used by the employees. Hence, it is challenging to identify the organisation’s exposure level and security posture.

The organisations’ perimetric walls are more able to block access to internal valuable assets and to protect them from outside intruders. Technology populism exposes the internal system to numerous unsecure connections that can be easily exploited by criminals to get privileged access, such as using public information sharing platforms or infected laptops or smartphones. Employees’ smartphones can easily move between domains of different security levels bridging the gap provisioned by a non-extended perimeter. Contractors and vendors, in their turn, bring their devices and machines to perform on-field maintenance operations and they have therefore to be connected to the internal network.

Mobile communication and computing devices include portable media (such as USB stick, CD, DVD, flash memory, or portable hard drive), laptops, tablets and smartphones. These devices, with or without internet or wireless connections, have the ability to connect or bridge two different networks. For example, if a laptop can be configured as an open access point, can readily and spontaneously provide a connection between an internal corporate network and a remote infected host. Mobile devices, even without wireless communication, can provide unchecked paths since they can move between zones of different security levels such as home and corporate networks. These different means have enabled not only legitimate user but also adversaries to remotely access assets and resources and to easily circumvent the deployed security perimeter.

3.7.3 Outsourcing

Lack of skilled personnel and limited resources capabilities as well as high costs for operations and maintenance are all driving forces of leveraging outsourced functions and resources. Consumers and end users increasingly demand online services and ubiquitous access to their data and transactions. Thus, in order to ensure adequate level of performance and efficiency, many organizations and critical infrastructure operators decide to outsource part of their business functions. Outsourcing some functions to external partners will allow further focusing on core assets and valuable resources. Thereby, the system management process will be performed in a more effective and efficient manner. Furthermore, outsourcing will play an important role in cost reduction process and in reaching skilled and competitive workforce over the wide world. It also enables flexible and on-demand usage of up-to-date technology with a minimum set of control, management and maintenance activities. As a result, organisations and critical infrastructure operators are increasingly engaged with outsourcing service providers to cost-effectively ensure agility and flexibility of their systems and to increase responsiveness to steadily changing business and environment conditions as well as increased demand on their services. To ensure quality of service delivery and business continuity, the
external outsourcing partners are usually provisioned with an access to the internal organizations infrastructure network. Third party access, on contrary, will open the door for more weaknesses and vulnerability providing potential intruders with unprecedented access to other core and valuable assets. In this regard, the main risk stems from the lack of control over providers or even over outsourced resources and the potential discrepancy between priorities of outsourcing service providers and their clients (e.g. critical infrastructure systems). This mismatch or disharmony between both sides can easily result in exposing critical business assets and resources to unauthorized access that could impact confidentiality, integrity and availability of the resources and hence safety, reliability and availability of the whole industrial control systems. Snowden disclosures, for example, have revealed that National Security Agency had access to data stored at some American cloud-based services and servers, such as Google and Yahoo [15]. The NSA spied on users collecting a lot of emails, contact lists, and search content, as well as tracking and mapping locations via mobile phones. Predominantly, encryption is used to create a protective perimeter around valuable and outsourced data. In this case, data cannot be separated from its perimeter. Therefore, any owned or collected data, which is stored, replicated to, or processed on off-premise resources, exchanged with other parties, even if it is encrypted is part of the organisation’s extended perimeter.

3.7.4 Human factor

The scale of a critical infrastructure system and the increased adoption of national and international collaboration and partnership models justify the constant tendency of such systems to extend beyond their conventional physical existence to include other entities such as vendors, business partners, service providers or even customers. Consequently, it is now a very common practice to see different external entities within the systems’ complex such as temporary workers, interns, independent contractors and subcontractors, or even visitors. Even if the access to the industrial sensitive zones is tightly controlled at the borders, behind the borders the freedom of movement is most likely ensured for ordinary organisation’s personnel as well as for temporary ones. Potential adversaries can exploit the dynamic nature of the systems as well as the lack of proper human resources management strategy to cause damage to the system. In the Stuxnet case, for example, outsider contracted programmer was infected first and then became unknowing carrier for the malware. Inadvertently, (s)he infected the air-gapped systems by transferring data from a personal laptop and flash drive [16].

Basically, human resources cannot be consistently confined behind static and predefined perimeter. Moreover, entities, such as employees, can also exploit their knowledge and privileged access to open paths of communication between inside and outside bypassing the perimeter infrastructure. Lack of risk awareness of employees can lead to numerous disastrous results. Furthermore, persons and their desires and motivations are the most difficult components to be constrained and with increasing rate of connectivity, the boundaries between individuals’ perimeter and corporate perimeter tends to vanish.

3.7.5 Summary

In a nutshell, Figure 1 shows how the organisations’ perimeter surrounds the assets and resources that embody values for the system to be protected creating so-called “internal world”. The non-extended part of the perimeter represents the wall, which maintain the segregation with the surrounding hostile environment (e.g. via fences, physical segregation of devices, virtual isolated domain), as well as doors, which provision a controlled and monitored access to the internal world for the legitimate entities (e.g. firewall, access controlled gate, network access control gateway). Moreover, the non-extended perimeter is usually supplemented by some monitoring and surveillance systems that provide twofold security enhancement, by increasing the deterrent effect of the perimeter and thereby discouraging an attacker from committing
undesired activities and crimes as well as maintaining a wide area situation awareness to boost the system responsiveness. Broadly speaking, the non-extended perimeter is responsible for controlling and providing authorized communication paths (green arrows in Figure 1) with the internal protected systems. As opposed to this, the extended perimeter part would create holes (red lines) in the respective defending wall impairing the system’s security posture and giving adversaries the opportunity to bypass the traditional security measures and hence to infiltrate and compromise the internal world. Therefore, it is important to maintain situational awareness within organisations so that the potential intruders can still be detected. Furthermore, the surveillance mechanisms investigated in this report take into account the dynamic nature of the utility networks and the limited available resources (e.g., security guards and badge check devices).

Figure 1 Overview of various components of an extended and non-extended perimeter
4 SURVEILLANCE TECHNOLOGIES

4.1 Overview of Surveillance Systems

The primary purpose of surveillance technologies is the collection of information about individuals, their activities, or their associates. This is achieved through the investigation or monitoring of the actions or communications of one or more individuals in certain areas. Currently, a wide range of surveillance technologies is used in order to provide end-users with different levels of functionality. Hence, for us to identify the various surveillance technologies, we conducted a systematic literature review in the course of WP4 and the results have been reported in Deliverable D4.1 [17]. This review method is capable of providing a valid and comprehensive categorization of existing technologies, and also helps in overcoming the difficulty of assigning various technologies with homogeneous groups. Specifically, the analysis resulted in the identification of six main categories of surveillance technologies, namely, biometrics; dataveillance; visual surveillance; communication surveillance; location tracking, and ubiquitous surveillance. Brief information about each of these categories is provided in the Deliverable D4.1.

However, apart from their main objective, certain surveillance technologies can be used as a mean towards the mitigation of various threats. Specifically, in [17] we examined the application of threat modelling techniques in utility networks and also investigated the effect of surveillance technologies on them. The following four sections provide a comprehensive overview of novel surveillance technologies, which have been examined and investigated in WP4 to enhance security and situational awareness. The main focus has been laid on Mobile ID Check technology and video surveillance as explained in Section 6.

4.2 Mobile ID Check

4.2.1 Motivation

Fast and secure identification of individuals is one of the keys to enforce security. Just showing a badge with some printed information is a weak strategy for access control. Developments for passport control in the field of border crossings shows a need for biometric verification to prove identity of an individual. This technique is rather good developed for stationary systems (see E-Gates for ABC – Automated Border Control), but for mobile systems, biometric technologies lack in terms of usability and secure devices. In HyRiM, we try to demonstrate a highly secure way to verify the identity of a person using a specialized device and a simple application prototype addressing the special needs in critical infrastructure, such as power plants.

4.2.2 Secure Mobile Identification Check

A risk management module evaluates the risk (e.g. depending on the area where the unusual event took place) and checks on the necessity to dispatch a security guard to the specific area or not. The guard makes his way to the location, and reports. If any person is present, the guard asks for identification credentials. The identification badge is scanned (via the camera of a special handheld device; special badges (OCR-B Text) can be easily printed on a simple sheet of paper for a demo) and the device sends the information to a server. Subsequently, the personal information (i.e., facial image) is sent back to the device (the person should be known to the system, either as an employee or as a visitor). The guard compares the facial biometrics information of the person to the claimed identity in the database (either manually or by help of the device) and acts accordingly. This biometric verification can also be done by fingerprint recognition but for the sake of simplicity we want to use the facial information which is easier to use.
4.2.3 Used Secure Elements (TPM and Secure Boot)

The following elements are used in the processing chain to make sure that biometric data is kept secure and to secure the device in a way that in case of stolen devices no reverse engineering or spoofing of applications or data is possible.

Devices for biometric data acquisition needs to have maximum protection against infiltration. To keep things secure, some secure elements should be introduced: The integrity of the device is ensured by a trusted platform module (TPM) [18] [19], beginning with a root-of-trust [20]. It starts with a design where the user has the proof that the application is the same at the stage of delivery and the application or the operating system was not changed or manipulated. To ensure this, the device start (or boot process) works with a so-called secure boot mechanism. The firmware and the application are stored in the device’s flash memory and are hashed or encrypted. The decryption key or hash value is stored inside the device in a memory which is not addressable after programming. This means that when programmed once, no one (including the manufacturer or software implementer) can reprogram the key – if a TPM is used. This is made impossible with the use of electronical fuses, which are destroyed after programming: it is impossible to read out these keys or hash values. The only access to the key is the CPU itself the TPM which reads out and proves the keys or hash values and compared them to the actual from the flash memory.

A device with a TPM and with an adapted software builds a Trusted Computing Platform [18] [19]. Such a trusted platform cannot be used against the interests of the owner or administrator, the platform is secured against manipulation from unauthorised third parties.

4.1.4 Demo system elements

We will briefly describe the demonstration system in Figure 2, which consists of a server (containing the reference facial images and the badge numbers as their index), a mobile device acting as a trusted computing platform and a person to identify. For demo purposes, the server can be a simple PC for the demonstration, for the real environment typical standard access mechanism would be applied and are sufficient. For the badge, a simple printed paper is sufficient and printed OCR-B text for the persons’ ID should be used for good scanning performance.

![Figure 2 System elements: Server with access permissions of persons, Secure ID check device](image)

4.1.5 Demo system workflow
At enrollment time (see Figure 3), an employee or visitor has to be inserted into the database with the following fields: Name, spatial-temporal access (e.g., Lab1 from 9am-19pm, whole area 24h), and a facial image. A simple badge is printed with a reference number to this person. The visitor has to show his badge at checkpoints or when checked manually. At verification time (see Figure 4), the device scans the badge with the reference number, asks the server for access permission and the facial image for this person and shows the guard (the person operating the check device) the relevant data. The device supports the guard indecision making.

1. Make a facial image

2. Edit access permissions

<table>
<thead>
<tr>
<th>Name</th>
<th>Given ID Nr.</th>
<th>Facial Image</th>
<th>Access spatial</th>
<th>Access temporal</th>
</tr>
</thead>
<tbody>
<tr>
<td>name 1</td>
<td>1</td>
<td></td>
<td>Lab1</td>
<td>9-19h</td>
</tr>
<tr>
<td>name 2</td>
<td>2</td>
<td></td>
<td>All</td>
<td>0-24h</td>
</tr>
<tr>
<td>name 3</td>
<td>3</td>
<td></td>
<td>Visitors platform</td>
<td>13:00-15:00</td>
</tr>
</tbody>
</table>

3. Print badge

**Figure 3 Enrollment of person**

1. Ask person for badge

2. Scan badge

3. System communicates with server’s database

4. Access permissions and biometrics shown and proved

**Figure 4 Verification of person**
4.3 3D Surveillance Camera

4.3.1 Motivation

Discovering frequent and rare spatio-temporal patterns in large amounts of streaming visual data is of great practical interest since it allows for automated applications of activity and surveillance analysis. However, the observed intensity pattern variations are often substantially influenced by dynamic lighting conditions such as shadows, highlights, reflections. Therefore it is required to have algorithmic methods and sensing modalities of high photometric invariance. Stereo vision generated dense 3D depth data is highly photometric invariant, and at the same time representing the geometry of a scene and its actors with a great accuracy. Due to these complementarities, we employ a combination of intensity (grayscale image sequence) and depth data to derive a highly specific (and simultaneously invariant) representation of spatio-temporal patterns.

4.3.2 Passive stereo depth sensing

Stereo vision is able to deliver dense 3D data in real-time and it can cope with a wide range of ambient light conditions. 2D image information is additionally available for free. Common camera hardware is reasonably priced, there are no moving parts, and camera housings can be made very robust. Disadvantages of stereo vision are that depth resolution decays with increasing distance and un-textured areas contain no depth information. However, the resolution decay is an issue that must be handled by algorithms that process and analyse the 3D data. Certain subjects such as individual persons, human crowd, clutter, textured floor, concrete, rocks and vegetation have a rich texture, which renders passive stereo working well for such type of subjects and environments.

![Figure 5 Stereo sensor heads of 40 cm baseline (left) and 20 cm baseline (right)](image)

We use an in-house developed sensor (see Figure 5) to extract intensity information, employing a canonical stereo setup (three monochrome cameras mounted in parallel), with different baselines (distances) between the two cameras located at the ends of the rig. The stereo matching process outputs depth data alongside with rectified intensity images, congruent to the depth image. Depth information is computed via a pyramidal implementation of a Census-based stereo matching algorithm, which is an explicit adaption and optimization of the well-known Census transform in respect to embedded real-time systems in software. Depth is computed for all three available baselines thus improving the quality of obtained depth map at the different spatial ranges. At the given resolution (of 1150×920 pixels), the sensor delivers approximately 10fps, when stereo computation is performed on a modern PC. Using the 40 cm stereo baseline (left image of Figure 5) with a spatial resolution of 1150×920 pixels, the computed depth data can be effectively used up to a distance of about 20 meters from the camera.

The main motivation behind the proposed multi-modal (depth + intensity) processing scheme stems from the complex outdoor detection task where typically strong variations in the observation conditions are
present. Multiple visual inputs of complementary nature are sought to improve weak or ambiguous detection responses while suppressing noise.

4.3.3 Algorithmic framework

Analyzing and recognizing spatio-temporal patterns in streaming visual data is of great practical interest given the recent explosive growth in the quantities of networked digital video. Many applied domains such as visual surveillance, ambient assisted living and activity-oriented video analysis seek to learn levels of normality and distinguish between frequently and rarely observed spatio-temporal patterns. However, the analysis task encompasses several challenges. Video streams exhibit a vast richness of information due to the inherent variability in the data thus large data amounts are required to obtain meaningful statistical models. Large data quantities are associated with a substantial computational cost and large memory footprint. Furthermore, structuring and modelling the data distribution can easily become nontrivial since data typically reside in high-dimensional feature spaces. The temporal characteristic of streaming data represents another challenge calling for computational techniques capable to build statistical models in an incremental and adaptive manner.

Two primary challenges in the context of rare event detection are represented by the following two issues:

The vast amount of data of the spatio-temporal image space imposes a challenge since its transmission and computational analysis call for substantial computational power and memory footprint. Another challenge is represented by the computationally efficient incremental clustering or vector quantization task in high-dimensional feature spaces.

Recent scientific research demonstrated that the structure of local, highly variable space-time patterns can be efficiently represented as a sparse linear combination of basis vectors using an over-complete dictionary. Recent work [21] [22] demonstrate that this finding can be used to learn a dictionary (basis vectors) representing frequently observed patterns (i.e. levels of normality) and a local space-time reconstruction step provides hints whether the local signal has been observed before (normal) or it represents an unusual situation (e.g. movement with a velocity exceeding normal, or previously unobserved presence at a location). The inclusion of 3D depth information results in making the inference process robust, since the enhanced photometric invariance and scale information associated with depth data can be well exploited. An incremental learning scheme implies that the dictionary is continuously updated and events which are first classified as unusual can turn over time into normal events if they occur frequently.
How to Enhance Perimeter Security Using New Surveillance Technologies

Figure 6 Computation scheme of the employed sparse spatio-temporal features $D = (D_1, \ldots, D_q)$, forming a local structural representation. Red cuboids illustrate some sample sparse structural units which can be used to represent the local spatio-temporal gradient structure.

Incremental robust sparse coding for representing local spatio-temporal signatures:

The main motivation behind using sparsity as the governing rule for creating an informative representation is the observation, that local spatio-temporal patterns carving out various shapes in the space-time feature space (see Figure 6) are very much self-similar, thus lie in a low-dimensional subspace. Hence, it is possible to find a set of basis functions (atomic representations defining a codebook) spanning the subspace, which can efficiently represent all observed variations of the spatio-temporal signal. Sparse coding, or in other terms, sparsity is a general constraint leading to model event patterns as a linear combination of a set of basis functions.

In mathematical terms, we can learn a dictionary $D \in \mathbb{R}^{p \times q}$ from a set of training features $[x_1, \ldots, x_n]$ which represent local signals from the streaming video:

$$
\min_\beta \|x - D\beta\|_2^2 \text{ subject to } \|\beta\|_0 \leq s
$$

where $\beta \in \mathbb{R}^{q \times 1}$ represents a set of sparse coefficients, which “activate” the learned dictionary entries within $D$ to best approximate the signal $x$. Therefore $\|x - D\beta\|_2^2$ is denoted as the data fitting term, where $\|\beta\|_0$ is the so-called sparsity-regulation term. Sparsity (the number of learned atomic components or basis functions) can be controlled by the sparsity parameter $s$, which is typically much smaller than the number of $q$, the number of signal dimensions.

At evaluation time we target to find $\beta$ based on previously learned set of codebook entry combinations - a combination that minimizes the reconstruction error of the locally observed signal. During the training time, frequently observed local signals induce the learned sparse dictionary, previously unobserved or rare signals.
will generate a large reconstruction error. This error, in least squares sense, is an indicator for the reconstruction quality and thus for the nature (usual versus outlier) of the local signal. This classification mechanism is depicted in the Figure 7.

**Figure 7** Illustration showing the testing stage of a previously unseen input signal. A set of (at training time) learned dictionary entry combinations are employed to test the reconstruction quality of the input signal. If the best reconstruction from $K$ combinations delivers a low error, it indicated that the input signal is well represented and it can be well approximated by the learned sparse code combinations. Otherwise, high reconstruction error indicates that similar signals have not been encoded in the testing combinations, and signals an unusual input signal.

The implementation of the proposed learning and evaluation framework has been carried out in the following manner:

- **Data**: Spatio-temporal image gradients of time-aggregated intensity and depth data are computed (after re-sampling images to a fixed smaller resolution (320x240 pixels)). The spatio-temporal volume is discretized to $M[pixels] \times M[pixels] \times T[frames]$ non-overlapping cuboid segments. Gradient signals from the intensity and the depth cuboids are concatenated to form a fused signal. Training is performed individually for each cuboid location, thus learning a location-specific signature for the observed time-varying signal.

- **Combination of intensity and depth data**: Spatio-temporal gradient features form the base for the features to be encoded. Intensity gradients represent finer details, but at the same time corrupted with photometrically induced structures. Depth gradients on the other hand provide contour information relatively free from illumination-effects, but exhibit (i) spatial noise (irregular boundaries), (ii) temporal noise (oscillating depth values) and (iii) invalid pixels, where no depth information is available.

- **Method**: at training time we learn at each location dominating combinations which will reconstruct the various signals encountered there over time. During test time, as shown in Figure 7, the objective is not to reconstruct the input signal, but to check whether one of the learned dictionary entry combinations is successful in reconstructing it. A small reconstruction error by one of the combinations is an indication of previously seen spatio-temporal patterns at the given location. Due
to the spatio-temporal discretization of the feature space outliers (unusual patterns) can be highlighted and coarse segments in the original video.

4.3.4 Evaluation plan and deployment

We recorded three staged scenarios by our stereo setup in a room and in a hall, with duration of 3 x 30 minutes. The usual pattern of pedestrian movement was walking (mingling) and waiting in a line, and we defined a set of unusual events which were performed at certain time instances of the recording by selected individuals. The table and the screenshot, as described in Figure 8, give an illustration of created scenarios.

<table>
<thead>
<tr>
<th>Abnormal Event</th>
<th>Visual Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loitering</td>
<td>Loitering off queue</td>
</tr>
<tr>
<td>Loitering</td>
<td>Loitering off queue</td>
</tr>
<tr>
<td>suspicious path</td>
<td>Joins queue then comes back</td>
</tr>
<tr>
<td>suspicious path</td>
<td>Abandons queue in the middle</td>
</tr>
<tr>
<td>suspicious path</td>
<td>Joins queue in the middle</td>
</tr>
<tr>
<td>Loitering</td>
<td>Loitering near queue</td>
</tr>
<tr>
<td>suspicious path</td>
<td>Joins queue in the middle</td>
</tr>
<tr>
<td>suspicious path</td>
<td>exits queue wrong direction</td>
</tr>
</tbody>
</table>

Figure 8 Left: table listing the staged unusual events for a selected indoor scenario. Right: a screenshot of the staged queue (normal behavior) and a loitering scenario (red bounding box annotation), and its equivalent top-view representation (inset B).

In each of the current scenarios we use 15 minutes for training (containing only normal, re-occurring situations) and 15 minutes for testing (also containing the occasional unusual events). The current staged scenarios serve the purpose of algorithmic development. We will record additional outdoor scenarios with more challenging viewing conditions and event structure for verifying the generalization capability of the employed sparse coding based technique.

4.4 ETRAFIC

Etrafi is a computer vision system capable of analysing video from fixed network cameras and providing results in the field of traffic control, such as queue length and duration, vehicle speed, lane occupancy and presence of stopped vehicles.

Figure 9 shows the main blocks in a typical video system, where several IP cameras are attached to a network with a video recorder and a monitoring system.
In the former diagram, each camera provides a video stream (either unicast or multicast) which is multiplexed in a network shared by all cameras. A video recorder or a monitoring system can subscribe to one or more video streams in order to get and manage the proper images.

In the same way, Etrafic can subscribe itself to a selection of video streams and analyze them in real-time, without interfering with any other system making use of the cameras (see Figure 10).

Since Etrafic system has been developed under the urban traffic management field, it provides two main functionalities: sensoring the road in order to obtain traffic management-related data, and detecting incidents on the road.

### 4.4.1 Traffic Management-related data

#### 4.4.1.1 Virtual loops

Virtual loops (see Figure 11) give information about the vehicles driving through the loop area:

- Lane intensity (vehicles per hour)
- Vehicle speed
- Time gap between successive vehicles
4.4.1.2 Queues

Queue sensor (see Figure 12) is able to measure parameters of a queue of stopped vehicles:

- Queue length
- Queue duration

4.4.1.3 Lane occupation

Lane occupation (see Figure 13) sensors can measure the following parameters:

- Spatial occupation of moving vehicles (%)
- Spatial occupation of stopped vehicles (%)

Figure 11 Virtual loop

Figure 12 Queues
4.4.2 Incident detection

4.4.2.1 Excessive queue

Traffic can be configured in order to provide an alert whenever one of the queues exceeds a particular length. Incident detection starts when a measured queue length goes above the configured activation threshold, and ends when a new detected queue keeps below the deactivation threshold. The definition of two different thresholds allows the alert to have hysteresis (see Figure 14).

4.4.2.2 Stopped vehicle

Vehicle stopped sensors detect those vehicles making illegal or emergency stops in those areas of the image that have been configured to be analyzed. An alert is triggered when a vehicle stops longer than the configured time threshold (see Figure 15).
4.4.2.3 Vehicle with emergency lights

This kind of sensor allows, in combination with the previous one, to detect those vehicles performing an emergency stop. This sensor analyses the video looking for intermittent lights, and its results are combined with those of the stopped vehicle sensor to trigger an alert whenever a stopped vehicle with emergency lights is detected (see Figure 16).

4.4.2.4 Kamikaze

Etrafic is able to trigger an alert whenever a vehicle is detected driving in the opposite direction of the lane.

4.4.2.5 Excess/Defect of traffic intensity

Etrafic is able to trigger an alert when the detected intensity on a lane or a group of lanes goes above the configured activation threshold. This alert is kept active until the intensity goes below the configured deactivation threshold. Similar approach is taken for the defect of traffic intensity alert (see Figure 17).
4.4.2.6 Camera failure

Etrafic is able to provide an alert whenever a video stream of a camera cannot be accessed.

4.4.3 Operation mode

Etrafic system can be operated both in standalone mode, or integrated with a control system which can receive all measured data and incidents and process them accordingly (see Figure 18).

In the case of standalone operation, Etrafic shows all information in a web interface with 4 main sections:

- Cameras: shows real-time video streams with the corresponding metrics printed on it. Real-time alerts are also shown in a sidebar.
- Measured data: provides access to historical data measured by Etrafic.
- Incidents: provides access to historical incidents detected by Etrafic.
- Configuration: allows the configuration of the different scenarios (video streams).
For integration with external control systems, a protocol built over TCP/IP has been defined.

### 4.4.4 Application in surveillance security

Since Etrafic is able of detecting stopped vehicles in preconfigured areas of the image, it can be configured as a non-invasive presence sensor in order to monitor and automatically trigger alerts about the behavior of vehicles in the perimeter or in the corridors of a controlled area. Usage examples include:

- Detecting unusual traffic flows at certain vehicle corridors (greater or lower than expected). Data provided by Etrafic can further be processed to compute more complex alerts
- Detecting presence of vehicles at certain sensible areas
- Detecting vehicles driving in the wrong direction or with unusual speed

### 4.5 Alert Monitor

In order to get alert notifications fast and in an easily interpretable manner, sensors can be integrated with the ETRA I+D Alert Monitor, an interface available both to the system operators and the technology provider. The purpose of this interface is to provide clear real-time information of the events happening in the different parts of a system, thus allowing the corresponding parties to take required actions as fast as possible and in possession of clear evidences of the problem.

#### 4.5.1 Alert operator interface

The Alert Monitor has a web-based interface (see Figure 19). Once an operator logs into the system, a dashboard is presented showing the status of the different facilities being monitored. For each facility, the following information is shown:

- Description of the facility, including name, picture and details
- Max. alert severity during the last 24 hours (ranging 0 to 10). Text colour signalizes the severity of the alert

By selecting any of the facilities, a detailed view of the alerts is provided:
• Timeline of alerts over last 24 hours: a line chart shows the evolution of the alerts over the time, per sensor.
• Distribution of alerts over last 24 hours: a pie chart shows the amount of alerts triggered per sensor, evidencing which sensors have been more active (see Figure 20).

![Figure 20 Amount of alerts triggered during the last 24 hours](image)

• Details per sensor: for each of the sensors, the max. alert severity and the amount of triggered alerts is shown. In addition, by selecting the sensor a detailed list of all the received alerts is presented (see Figure 21). Alert operator can choose to ignore the alerts once they have been checked.

![Figure 21 Details per sensor](image)

### 4.5.2 Integration of sensors

The alert monitor provides different interfaces for sensor integration:

• REST API: any system capable of triggering alerts can benefit from the REST API to push its alerts into the Alert Monitor, just by sending an HTTP POST request with the details of the alert.
• Legacy systems: a framework has been developed in order to allow a seamless integration of “legacy” systems with the Alert Monitor, by directly processing their log files. Whenever a certain line (defined with a regex) is found, the configured alert logic is triggered. This alert logic can be defined in order to:
  o Trigger an alarm whenever a certain log line is found a number of times (1-N) in a certain time range.
  o Certain combinations of log lines are found within a certain time range

4.5.3 Application in surveillance security

The purpose of the Alert Monitor is offering to the security operators an aggregated view of all the events and alerts being triggered by different systems monitoring their facilities under control. No limits are imposed to the nature of the systems, so both physical and cyber-security related sensors can be integrated with minor works with the Alert Monitor, as far as they produce log files or are capable of sending HTTP requests. Security operators can take benefit of the features of the Alert Monitor:
• A dashboard showing summarized information about the situation of a facility, being calculated with information coming from a variety of sources
• Access to alerts provided by different sensors from a unique site, where temporal correlation between the alerts being triggered by independent sensors can reveal further information about the problem being assessed
• Centralized access to the details of the alerts being triggered by independent sensors

5 CHALLENGES OF SURVEILLANCE TECHNOLOGIES

In this section, we discuss the problems when dealing with security and privacy in the surveillance systems. Our considerations start with an analysis of security and privacy threats. And then, corresponding data protection acts in EU and available protection technologies are reviewed. Finally, the most relevant challenges with the involved tradeoffs are discussed.

5.1 Security and privacy threats in surveillance systems

Surveillance systems have possible threats on both servers and security devices. Those threats can be classified based on their goals. In this subsection, we adopt the attack tree modeling method to analyze potential security and privacy threats. As described in Figure 22, threats can be classified as privacy threats (Pa) and security threats (Pb) in surveillance systems. Privacy threats can be roughly categorized as inside attackers (P1) and outside attacks (P2), which can be further categorized as interception (P2.1) and information exfiltration (P2.2). At the same time, security threats can be roughly categorized as modification (P3) and fabrication (P4).
Table 1 Transitions in the threats model

<table>
<thead>
<tr>
<th>Notation</th>
<th>Transitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1:</td>
<td>Data leakage to outside</td>
</tr>
<tr>
<td>T2:</td>
<td>Privacy leakage</td>
</tr>
<tr>
<td>T3:</td>
<td>Misinforming entities in the systems</td>
</tr>
<tr>
<td>T4:</td>
<td>Raising risks in the systems</td>
</tr>
</tbody>
</table>

Those leaf attacks are described as following:

**P2.1 Interception**

Surveillance communication is the act of capturing surveillance data as it travels across the wireless network towards the server. The units being monitored are often referred to as packets. Packets are the broken up parts of the surveillance data e.g., finger, images and video over Internet protocols. The adversary usually has powerful computational information eavesdropping devices to intercept packets and open up the processing surveillance information. No matter the aim of the surveillance is to detect suspicious conditions of human or inanimate objects, those data are with the targets’ location information. If the information is also with personal information, e.g., a video camera photographs a person passing by an aisle. The target would not like to share this information because it will affect its daily life. The goal of this attacker is to use the information for its own purposes such as scouting a certain region of interest while remaining undetected.
P2.2 Information Exfiltration
Besides communication, target’s biometric data also should be protected during data access, since tons of biometrics data are stored in servers. Data access issue is mainly related to security policies provided to the users while accessing the data. The server’s security and privacy requirements, including authority and confidentiality, affect the functionality of the surveillance system. If the adversary directly gets data from the server, targets’ privacy will be leaked.

P1 Inside Attackers
Different from other attackers, employees have legitimate access to the device control and data access, e.g., system operators or security guards. Those adversaries can perform the same attacks as outsiders and therefore the same security requirements apply. Corresponding laws must be put in place to ensure that only minimal required amount of information is disclosed. Moreover, technical precautions must be taken such that unauthorized entities should not have the right to access the corresponding usage data.

P3 Modification
The communication between servers and surveillance devices is usually based on a wireless network. As a kind of network, it has the same vulnerability as other wireless networks. In this scenario, the attackers seek to make the server unavailable to the surveillance devices by temporarily disrupting services, e.g. denial of service attacks. Therefore, the server waste resources to these alarms and neglect other attacks.

P4 Fabrication
Not only servers, but also security guards are the aim of attacks. The attackers launch attacks to modify a legitimate message or to frame a server that would affect the security guard’s decision, e.g. spoofing attack, message replay attack and man-in-the-middle attack. This attack directly violate physical access controls in surveillance systems.

5.2 Legal and Regulation Aspects
Privacy in surveillance systems has attracted extensive attention from government, industries and academia due to their importance. It is an important consideration of human right law, especially in EU countries. The European Convention on Human Rights state rights to respect for private family life:

1. Everyone has the right to respect for his private and family life, his home and his correspondence.

2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

Therefore, it is necessary to explore data protection laws of EU and its countries [23, 24, 25, 26, 27, 28, 29]. The HyRIM consortium comprises 7 partners from 5 different EU countries, including Austrian, Germany, Italy, Spain and UK. In the end, the on-demand surveillance system will be experimented in Akhela, Italy. Therefore, Akhela code of ethics should also be considered. In this subsection, target’s biometrics data protection will be depicted.

General Data Protection Regulation (GDPR)
Replacing EU Data Protection Directive (Directive 95/46/EC), the GDPR is directly applicable in all member states without the need for implementing national legislation. The GDPR focuses on EU citizen’s privacy
(Article 4), places onerous accountability obligations on data controllers (Article 5), lays down data processing (Article 6-11) and defines data subjects’ rights (Chapter 3).

**Austrian Federal Act Concerning the Protection of Personal Data (DSG 2000)**

DSG 2000 does not specify privacy in surveillance systems, but it refers the use of sensitive data and also defines processing requirements (Section 10). The controller of data application shall inform the data subjects when collecting data in an appropriate manner about:

1. The purpose of the data application for which the data are collected;
2. The name and address of the controller.

Insofar as this information is not already available to the data subject, with regard to the particular circumstances of the case.

**Germany Federal Data Protection Act (BDSG)**

BDSG specifies monitoring of public accessible areas with optic-electronic devices (Section 6b). According to the item of the law,

1. It is only allowable to fulfill public tasks; to exercise the right to determine the access; or to pursue rightful interests;
2. The fact that the area is being monitored and the controller’s identity shall be made discernible by appropriate means;
3. The subject of the biometrics data should be informed of the storage;
4. Biometrics data should be deleted without delay, if they are no longer needed for the pursued purpose.

**Italian Personal Data Protection Code**

The data protection code also mentions surveillance technologies and biometrics data (Section 56). If the biometrics data should be processed, measures and precautions should be made up to safeguard data subjects. Personal data processing should under agreement of subjects. The communication of data processing should be guaranteed by digital signatures or by facsimile or registered letter. The processed data should be erased in 45 days after the receipt of the data (Section 39).

**Spanish Organic Law on the Personal Data Protection**

The controller shall be obliged to implement the right of rectification or cancellation of the data subject within a period of ten days. The provisions on the deletion of files shall lay down the fate of the files or, where applicable, the timetables to be adopted for their destruction.

**UK Data Protection Act 1998 (DPA)**

Although DPA does not mention biometric data protection, it was enacted to bring British law in line with the EU Data Protection Directive which required member states to protect people’s fundamental rights and freedoms and in particular their right to privacy with respect to the processing of personal data. DPA defines rights and responsibility of data subjects, controllers, enforcement notice and exemptions.

**Akhela Code of Ethics (ACE)**

As one of the most important partners, Akhela provides experimental environment for the on-demand surveillance system. Therefore, Akhela code of ethics has to be followed. According to ACE, the confidentiality
of the processing data should be ensured, strictly refraining from mining confidential data through illegal means (Subsection 4.4). Akhela should comply with all statutory provisions governing the privacy of employees and collaborators. However, Akhela employees and collaborators should exercise due care and diligence to protect corporate assets (Subsection 5.4).

Privacy is surveillant subjects’ first concern, but it is often overlooked. Legal decisions on privacy are expectation of privacy. It is difficult to judge, especially in current surveillance systems. According to the European Convention on Human Rights, Article 8 (Right to respect for private and family life) and Article 10 (Right to freedom of expression) should be balanced. Moreover, those judgement depends on corresponding technologies. Therefore, some technical solutions in surveillance systems are summarized in the next subsection.

5.3 Technical Solutions

Legal approaches need supports from technical aspects. In this subsection, we summarize security and privacy technical countermeasures in surveillance systems.

Human Motion Recognition

Human motion recognition is now in widespread study. For surveillance systems, human motion recognition is an important approach to detect physical intrusion. Motion recognition is firstly proposed by Stauffer and Grimson [30]. The proposed approach presents the idea of representing each pixel by a mixture of Gaussians and updating each pixel with new Gaussians during run-time. This allows background subtraction to be used in outdoor environments. It is a very important first step to recognize human motions. Sidenbladh [31] calculates optical flow for a large number of image windows each containing a walking human. A support vector machine (SVM) is used to detect walking humans in video. Optical flow can be noisy and instead image flow can be measured using higher level entities. More popular motion recognition approaches focus on visions. Yamato [32] utilizes low level silhouettes of human actions in a Hidden Markov Model framework, where binary silhouettes of background-subtracted images are vector quantized and used as input to the models.

Secure Processing

Secure processing is the basic security and privacy requirements of the surveillance systems. If a surveillance report is authentic, the system should recognize if it is copied or cloned. Modern encryption techniques, particularly PKI (Public Key Infrastructure) scheme [33] aims to enable the authorized entity to verify the authenticity and integrity of the data stored in the system. The specifications do not try to prescribe a full implementation of a complicated PKI structure, but rather are intended to provide a way of implementation in which states are able to make choices in several areas (such as active authentication, anti-skimming and access control, automated border crossing, etc.), thus having the possibility to phase in implementation of additional features without being incompatible with the total framework. Ito et al. [34] presented an image matching algorithm using phase-only correlation in the frequency domain. This algorithm preserves the visual privacy of the images in a template database. In order to achieve this, all the images of the template database are converted to the frequency domain. Then, a phase scrambling using a one-time key is applied to the discrete Fourier transformation coefficients. Afterwards, in order to match a query image with an image of the templates database, the query image is converted to the frequency domain and the matching is done with POC (Phase-Only Correlation) using the DFT (Discrete Fourier Transforms) coefficients as inputs. In order to ensure data integrity and identity authentication, more literatures focus on TPM (Trusted Platform Modular) with a "secure boot mechanism". Sirotich [35] utilizes RFID chip technology to ensure ePassport security. This means that the application and the operating system in the device cannot be changed. It is
encrypted or hashed and the key (or the access mechanisms to read out the key) for decrypting/ or checking the hash are destroyed after being programmed. Only the CPU itself inside the chip knows to access the data. So if the device is stolen, and if attackers can disassemble the code from the flash store, they cannot do anything with it. The CPU only accepts exact THIS loaded image.

**Role-based Access Control Model (RBAC)**

Role-based access control model is the most popular approach to control data access. In the on-demand surveillance systems, location is an important piece of information. Daniela et al. [36] introduces Areas of Responsibility (AOR) into RBAC, which describe the separation of huge-region-smart grids into smaller entities. Those AOR can in turn represent a logical set of smaller regional groups, if the regional groups consist of a unique set of elements in the smart grid. In order to provide dynamic access control in smart grid environment, the context awareness-based dynamic access control model requires a specific structure [37]. The proposed model utilizes context information, which is collected during the user authentication phase using a context awareness manager. It performs mapping and enforces the policies of the relevant database. During the access right decision phase, it also classifies the task, tags and applies the role in context. User context information is determined by the user’s environment during access request.

**5.4 Open Questions**

Government and academia have expressed concerns regarding security and privacy in surveillance systems. By analyzing the state of the art of security and privacy in surveillance systems, open questions, e.g. threat detection and privacy evaluation, are listed in this subsection.

**Threat detection accuracy VS. Efficiency performance**

Threat detection is the most important aim of surveillance systems. The alert monitor should predict accurate risks. There is considerable difference of opinion as to what should be the different risk levels. Security guards can react according to those levels. However, higher quality detection accuracy requires more data and higher quality data. As we mentioned in the previous part, portability of the device is one of the most important features in the HyRiM surveillance system. That means the device would not be very heavy or bulky. This also limits the amount of processing power and available memory. However, surveillance data processing and information analysis are power consuming in the system. Only a little resource is left for privacy protection and security. Therefore, a suitable tradeoff between system performance and the implemented security functionalities is a practical challenge.

**Data subsect privacy VS. Surveillance awareness**

An important reason for privacy concern is that the surveillance data carries their subjects’ biometrics and location information. Using those data, the subjects’ can be easily tracked throughout their daily lives. Section 5.1 described that surveillance data can be eavesdropped during data transmission and storage. The popular wireless network infrastructure opens more vulnerabilities for privacy violation. It makes privacy preservation more challenging in surveillance systems. In current working environment, however, employees usually still have little knowledge on the surveillance systems and barely any influence on the collected information. Obviously, technical solutions are not sufficient to address all privacy problems on surveillance awareness due to diverging interests of employer and employees. Corresponding policies should be required to foster actual deployment of security and privacy solutions in HyRiM. On the one hand, employers should notify what kind of data is collected, how it is used, who has access to the data and how long the data will be stored; on the other, privacy metrics should be built to inform employers to avoid privacy breach with surveillance technologies.
6 HOW TO ENHANCE SECURITY USING SURVEILLANCE TECHNOLOGIES

6.1 Motivation (Limitations of Traditional Surveillance Systems)

Many infrastructures (e.g., supply networks, large-scale enterprises, etc.) spread over large geographic area and across long distances connecting regions, which are geo-graphically very far apart. Moreover, these organizations constantly tend to extend beyond their physical existence to include other entities such as vendors, business partners, service providers or even costumers. Thus, it is very common to have external entities inside their sites such as temporary workers, interns, independent contractors and subcontractors, or even visitors because of their extended perimeter (for more details see Section 3). The adversaries can exploit the fact that the issued badges, for example, of terminated employees or temporary visitors and workers are not always timely recovered before leaving the site and the access of stolen and lost badges are similarly not revoked in a timely manner. As a consequence, the perimeter-centric physical security measures such as traditional surveillance technologies (e.g., Closed Circuit Television (CCTV) systems or entry access control solutions) that use static surveillance devices mounted at specific locations are not adequate to detect and prevent such potential intruders [38]. Due to the inflexibility and fixed installation of these systems, their deterrent effect will be considerably less. The monitoring system predictability is very high and the attacker can observe and adapt to the current surveillance strategy. Therefore, it is important to maintain situational awareness within the industry complex so that the potential intruders can still be detected.

To cope with this dynamic nature and to achieve adequate level of situational awareness in such large-scale areas taking account of limited available resources (e.g., security guards and badge check devices) surveillance strategies have to exhibit two key features, namely risk-based and on-demand. Risk-based strategies are necessary to allocate and focus resources and capabilities in the high sensitive areas and against high threats, and therefore to effectively and efficiently mitigate risks [39, 40]. On-demand strategies, on the other hand, are required to randomize security checks to improve flexibility and detection probability and hence to enhance the organization’s security posture. This feature can be achieved by mobilizing the required security resources. Furthermore, the application of concepts from game theory will help in finding the optimal multi-goal surveillance strategy, which simultaneously optimizes different goals such as effectiveness (detection rate), privacy and costs.

6.2 Game-Theoretic Approach for Surveillance Enhancement

6.2.1 Standard Model

It is convenient to think of the utility provider’s infrastructure as a finite undirected graph $G = (V, E)$ with nodes in $V$ corresponding to physical places, and edges in $E$ being associated with connection paths between them. Without loss of generality, we may assume edges to be without surveillance, since we can always model any path (e.g., an aisle) under surveillance as another node in the middle of the edge. More formally, if zones A and B are connected by an aisle and that aisle is under surveillance (e.g., by a camera), then it is treated as a third place C with the graph model having the edge sequence $A \rightarrow C \rightarrow B$, instead of the single edge $A \rightarrow B$ in which the aisle would be assumed without any protection or detection mechanism. In this view, the intruder may (randomly) walk on the graph in an attempt to reach his goal (the zone with the valuable business assets) while avoiding meeting the security personnel at any node. In case the intruder is captured, it gets kicked out of the building (removed from the graph), and the game-play starts afresh again. Putting this in a more formal way, let a single pure strategy for the defender be a circle in the infrastructure graph $G$. That is, a surveillance tour always starts and ends at the same node (though not all tours may originate from the same place). The entire strategy space of the defender being the “surveillance staff” is thus a (not necessarily minimal) set of circles $C_1, \ldots, C_n$ that spans $G$. On the contrary, the attacker’s action
set be a set of paths $P_1, ..., P_m$ that all, w.l.o.g.\footnote{the case of multiple assets being under attack is easily covered by adding a hypothetical node $v_0 \not\in V$ to $G$ and connecting all potential target places to $v_0$, so that all attack paths end in $v_0$.}, end at a specific valuable target node $v_0 \in V$, and may start at different (arbitrary) points outside the infrastructure. Note that the setup is here slightly different from a conventional pursuit-evasion game \cite{41} \cite{42}, since in our case, it is not known whether or not there is an intruder in the network (this asymmetry is usually not found in a simple pursuit-evasion game, in which – by construction – the players know about each other’s existence). In the classical version of the pursuit-evasion game, the payoff in the game would correspond to the outcome of the detection of the intruder. In this case, the game itself becomes a simple matrix-game, whose payoffs are stochastic in the sense that the payoff matrix $A = (A_{ij})_{i,j=1}^{n,m}$ is one of the Bernoulli random variables $A_{ij} \sim Ber(p_{ij})$ with the semantic that:

$$A_{ij} =
\begin{cases}
0, & \text{if the intruder is missed;}
1, & \text{if the intruder is caught.}
\end{cases}
$$

in which the parameter $p_{ij}$ tells how likely a detection of the path $P_j$ along the tour $C_i$ is. Packing all temporal matters and detection errors into the simulation (as discussed in Section 6.4.1), it is an easy yet laborious matter of working out the specific distributions, i.e, their parameter(s) $p_{ij}$. Solutions in the sense of Nash-equilibria of the resulting “non-deterministic” game can be obtained in various ways. The most obvious one is to convert the matrix of random variables into a real-valued matrix by taking the expectation per element. This results in a real-valued matrix $B = (B_{ij})_{i,j=1}^{n,m}$ that can be treated with the entire well-known machinery of game-theory (von Neumann’s minimax theorem and linear optimization).

6.2.2 Game Theory using Uncertainty

The abstract and basic model sketched in the previous section can hardly be described accurately by crisp outcomes as in classical game theory (for the reasons mentioned above). Additionally, in real-world surveillance systems there are several practicalities and imperfections that can significantly result in a fluctuating detection performance of the system. For example, every surveillance camera system has blind spots, and not every person in an inspected zone may be caught or available for a quick automated identity check, though s/he may legitimately be there. Moreover, there are many unforeseen events in the context of mobile ID check scenario that may also force the security personnel to change their checking strategies, which leads to random detection performance/reliability. These are pieces of uncertainty that have to be reflected in the model. Additionally, we are more interested in the expected damage caused by the intruder as just detecting his/her presence. Thus, we will use the expected damage as the payoff of the game. In this context, the expected damage still relies on the likelihood of detecting an intruder. However, it is also important to consider in which area an intruder is currently located since, if the intruder goes undetected, this influences the magnitude of the expected damage. In other words, an intruder can cause more damage in a “high” security area than in a “low” security area since the assets are of different value for the organization.

We assume the payoff to be quantified not by a single number, but rather be described by a set of possible outcomes that either stem from simulations, surveys, or expert interviews. In any case, a real-valued payoff matrix, similar to matrix $B$ and based on the Bernoulli random variables from matrix $A$ in eq. (1) is no longer appropriate. And we need to resort to a more expressive categorical distribution to avoid information loss. Putting this in a more formal way, let a single pure strategy in the model be a set of frequencies $f = (f_1, ..., f_M, f_R)$ representing the amount of times a security guard is performing a security check (tour, resp. circle in the graph $G$) in a “low”, “medium” and “high” area, respectively. Hence, the strategy space is the collection $f_1 ... f_n$ of all these frequencies, at which the tours $C_1, ..., C_n$ are being taken. That is, we seek to optimize the frequency at which the guards take their turns to walk around the premises, and – in the most general setting – we could let each tour at each frequency constitute its own strategy in the game. Practically, matters will have to be simplified by restricting oneself to certain “reasonable” intervals of taking tours,
which makes the strategy set boil down to practically much less than the full set of combinations of tours and frequencies. Similarly, the adversary’s intentions may concern a wider set of security zones \( Z_1, Z_2, \ldots \), where it wants to cause some damage, and each zone could be reached on different possible paths. Again, for simplicity (and consistency of the presentation here), let us assume that there are \( m \) attack strategies \textit{in total}, which could – for example – correspond to a (sub)set of security zones.

The payoff matrix \( A = (A_{ij})_{i,j=1}^{n,m} \) for this game contains probability distributions as entries instead of single numbers, i.e.

\[
A_{ij} \sim F_{ij}(\text{dat}_{ij}).
\]

where \( F_{ij} \) is a probability distribution with density \( f_{ij} \), which is the kernel density estimation of the observed data \( \text{dat}_{ij} \) for the scenario \( ij \) (e.g., based on all available expert opinions assessing the expected damage, the performance of the surveillance system in terms of incident recognition and classification, etc.). This data vector \( \text{dat}_{ij} \) can be obtained from simulations, say to be a collection of indicator variables recording the event of the intruder having been detected, or not (in that case, \( \text{dat}_{ij} \) is a 0-1-valued array). In a more fine-grained view, we can let a simulation run and record in \( \text{dat}_{ij} \) whether an intruder has been gone undetected or has been caught; and if it has been caught, which area has it been. That is, we could note a low risk in the vector \( \text{dat}_{ij} \) if the intruder was caught in a low security area, or a high risk if the intruder made it into a high security part of the infrastructure.

To preserve all the information from the simulation, its entire output goes into a probability distribution \( \tilde{F}_{ij} \), so that we can invoke the more flexible framework developed in WP1 and put forth in [74] alternatively to the standard minimax and optimization approach. This allows us to play the game directly with the distributions rather than having to convert them into “representative” real numbers. Moreover, we can add several more dimensions to the game-play optimization, such as cost for a route, e.g., in terms of the time it takes to traverse round trip \( C_{ij} \), i.e., to go to a specific zone and perform the security checks therein, or of the inconvenience caused by unwanted and too frequent identity checks (since they might interrupt the current work of a person or might not be possible immediately). However, the most important benefit from directly working with the distribution is gained when the Bernoulli-distribution is replaced by a more general, categorical or even continuous, distribution model over the categorical damage scale that applies to the indicators mentioned above (e.g., detection rates, privacy infringement, comfort, etc.).

In general, the payoff structure remains random but is now a collection of categorical variables (measuring the risk, damage, inconvenience, etc.). Converting such a model into a real-valued model by taking expectations amounts to averaging ranks of the respective categories loses the semantic of the model and blurs much of the information obtained from the surveillance system. To avoid these issues, the framework developed in WP1 and presented in [74, 75, 76], which integrates uncertainty and distribution-valued payoffs into game theory, can be applied. In the following, we will briefly introduce the basic notions of the framework again for convenience of the reader and state how they relate to the aforementioned problem of surveillance.

Let the random variable \( X \) which can be continuous, discrete or categorical, represent a(ny) payoff in the matrix structure, and assume that \( X \) is supported on a compact set (and has a continuous probability density function in case that \( X \) is a continuous random variable). We represent \( X \) by the sequence of its moments, treating this sequence as a hyperreal number \( x = (E(X^n))_{n \in \mathbb{N}} \). It is an easy matter to verify that \( X \), and respectively its distribution function, is uniquely represented by the sequence of its moments, and that any two variables are \( \leq \)-ordered in the hyperreal space \( \mathbb{R}^\ast \). Transferring this ordering to random variables \( X_1, X_2 \) with distributions \( F_1, F_2 \), we write \( X_1 \leq X_2 \), resp. \( F_1 \leq F_2 \), if the corresponding hyperreal representatives are \( x_1 \leq x_2 \).
Under this embedding of distributions into \( \mathbb{R}_* \), we can play the game “as usual”, only bearing in mind that the gameplay itself is now over a new algebraic structure. Things are, however, greatly simplified in the sense that we do not have to deal with hyperreal arithmetic, based on the following facts (cf., [75] for proofs):

- Two distributions can be compared by looking at their tails. Specifically,
  - if the distributions are categorical, written as \( F_1 = (p_1, ..., p_n), F_2 = (q_1, ..., q_n) \), where both distributions are ordered along descending ranks, then \( F_1 \preceq F_2 \) if and only if the vector \( F_1 \) is less or equal to the vector \( F_2 \) in terms of the usual lexicographic ordering;
  - if the distributions are discrete or continuous (with compact support), then we can truncate them to become supported on a compact set. Truncated discrete distributions then compare as categorical distributions, and continuous distributions compare lexicographically under a slightly more complicated representation that we do not look at here (as being not required for our current application);
  - if one of the two distributions is degenerate, say \( X_1 = \alpha \) is a constant (a deterministic outcome) and \( X_2 \) is random and ranges within the set \( [x_1, x_2] \subset [1, \infty) \), then \( X_1 \preceq X_2 \) if and only if \( \alpha < x_2 \) (conversely, \( X_1 \preceq X_2 \) if and only if \( x_2 \leq \alpha \)).

In any case, the decision \( F_1 \preceq F_2 \) can be made without resorting to any hyperreal arithmetic.

- There are modified versions of fictitious play (FP) to solve zero-sum matrix games with probability distribution-valued payoffs (note that [75] gives an example demonstrating that regular FP like in [77] can fail to converge although the game is zero-sum).

It has to be pointed out that the special case of Bernoulli distributions is canonically covered by the framework just sketched, since the lexicographic ordering on this distribution (with only two categories) equals the natural ordering of the real-valued expectations. Thus, the simple approach of converting 0-1-valued random values into their averages for a game-theoretic treatment is an easy special case of the above framework.

To make an integration of the framework into software-based risk management tools easier, all of the functionalities of the framework have been implemented in R [Error! Reference source not found.]. Such a tool facilitates the handling of the game-theoretic algorithms for a risk manager, keeping away the burden of with data aggregation or consensus finding. Using the R package implementing the game-theoretic framework, a risk manager can safely rely on theory and algorithms to support his decisions purely based on all the available data. Section 6.3 and Section 6.4 illustrate the application of the aforementioned framework onto video surveillance and Mobile ID Check technology.

**Remark:** in a different application style of the optimization (game), one could take the frequencies of certain areas being checked as induced by the tours that the guards take. That is, an area may be checked several times during a certain tour \( C_i \), or only once. The latter induces a low frequency of checks for the area, while the former yields more frequent checks. The game-play for the defender would then be about optimizing the circles to achieve the best frequencies according to the respective security demand per area (constrained to a full coverage of the entire infrastructure). In this way, we could optimize the surveillance tour layout itself, but this is a different story and out of the scope here. In any case, such an approach may be prone to combinatorial explosion of strategies, so that we will stick with the more feasible optimization of the surveillance configuration in terms of optimal frequencies at fixed tours, rather than optimal tours for optimized frequencies.
6.3 Security Enhancement Using Video Surveillance Systems

6.3.1 Introduction

Video surveillance systems capture video streams from the surveilled areas. The retrieval and prediction of risk-related patterns from video stream data, combining with the HRMs developed in HyRIM, can be used to enhance security of the surveilled areas. However, retrieval and clustering of potential risk-related patterns from video camera surveillance systems is a challenging problem, as the object being tracked can undergo considerable changes (e.g., changes in scene illumination, presence of shadows, changes in the pose of the object, etc). Given a set of surveilled image data, we aim to extract/predict patterns and cluster those patterns by similarity. Based on these clusters and the requirement from the security operator of the surveillance areas, risk levels are assigned to each cluster. Finally, loss distributions as functions of risk levels will be compiled from the video surveillance data and the optimal strategies for the surveillance system to enhance security will be suggested.

6.3.2 Pattern Retrieval from surveilled Image Data

In order to compile risk levels from surveilled image data, the potential patterns which contain risk factors should be first retrieved. In this deliverable, Bag of Features (BoF) [43] and specifically Lowe’s 128-dimension Scale Invariant Feature Transform (SIFT) [44] descriptor is utilized to describe the local patch in images which are obtained from video surveillance systems. The basic idea of SIFT is to look for the extreme points in the scale space, filter these extreme points to find the keypoints and describe the keypoints by 128 dimensional vectors. The 128 dimensional SIFT descriptor is a histogram of responses to oriented gradient filters. SIFT transforms an image into a large collection of feature vectors, each of which is invariant to image translation, scaling, and rotation. SIFT has been empirically proven to be one of the most robust local invariant feature descriptors [45].

BoF methods have many application scenarios, including image classification, object detection, image retrieval, and so on. The most important characteristic of BoF approaches is the use of an orderless collection of image features. Recent research has demonstrated its effectiveness in image processing [46]. The BoF term vector is a compact representation of an image which discards large-scale spatial information, the relative locations, scales, and orientations of features. The image features represent local areas of the image, just as words are local features of a document. Feature detection is the process of deciding where and at what scale to sample an image. One general feature detection method is interest point operators [43]. The output of interest point operators is a set of keypoints (that are stable under minor affine and photometric transformations) that specify locations in the image with corresponding scales and orientations. The most popular keypoint detector is that developed by Lowe [44].

The location of SIFT feature is defined as extrema of the result of Difference of Gaussian (DoG) functions applied in a scale space to a series of smoothed and resampled images. A scale space of $I$ is generated by convolving the original image $I(x, y)$ with the Gaussian blur $G(x, y, k\sigma)$ at scale $\sigma$

$$L(x, y, \sigma) = G(x, y, k\sigma) * I(x, y)$$

Thus, a DoG image $D(x, y, \sigma)$ is given by

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$

where $k > 1$ and $(x, y)$ is the pixel location.

Once DoG images have been obtained, the location of the SIFT features are the local/minima/maxima of the DoG images across scales. Specifically, a pixel value is marked as maximum or minimum among all compared pixels: eight neighbors at the same scale and nine corresponding neighboring pixels in each of the neighboring scales.
Feature descriptor is used to represent the neighborhood of pixels near a localized region. The description of a SIFT feature is derived from the neighborhood in the blurred field of the corresponding scale. A 16 x 16 neighborhood around the keypoints is taken. For 2D SIFT features, the local image gradient uniquely defines the orientation of a 2D neighborhood [44]. The gradient of the blurred scalar field is sampled and a set of 2D orientation histogram is recorded. Each histogram, which records the magnitude of the gradient, is part of the SIFT descriptor. The descriptor then is a vector of all the values of these histogram and is normalized to have unit length.

In this deliverable, we design a pattern retrieval model by utilizing SIFT features, based on the bag-of-feature representations. Features are computed for each image to form a high dimensional (4x4x8 dimensional) descriptor. We consider each feature as a pattern, which can be further referred to as a visual word. Each image is then represented as a distribution on visual words. Figure 23 presents the establishment procedures of the BoF SIFT algorithm based pattern retrieval model. As shown in Figure 23, at first, the interest point will be detected and then keypoints are generated, which will be stored as a 128 dimensional vector.

![Pattern detection in images](image)

**Figure 23 Procedures of the BOF SIFT algorithm**

### 6.3.3 Pattern Clustering and Risk Level Assignment

All the visual words produce a codebook, which is also one key concept in BoF. K-means is one of the simplest unsupervised algorithm that has been widely used in image processing [47]. It is also widely used to cluster SIFT descriptors to form a codebook [48, 49]. K-means is a method of vector quantization. The quantization of visual words affects the image retrieval result strongly [49]. The most challenging problem of K-means that affects the effectiveness of pattern clustering is the selection of cluster centroids. In the HyRiM project, we suppose there are totally five risk levels, i.e., [very high, high, moderate, low, very low], which is corresponding to five clusters of patterns. In this deliverable, we consider an improved version of K-means, which is described in detail in Algorithm 1.
Algorithm 1: Improved K-means algorithm

1. Initialize $K = 5$ cluster centroids $\sigma_1, \sigma_2, \cdots, \sigma_K \in \mathbb{R}^{128}$ randomly
2. Repeat until convergence:
   { 
   3. 2.1 For every feature $x^{(i)} \in \mathbb{R}^{128}$, set the cluster it should belong to as $c^{(i)} = \arg \max_j |x^{(i)} - \sigma_j|^2$
   4. $\% c^{(i)}$ is the index (from 1 to K) of cluster centroid
   5. 2.2 For each cluster $j$, set the new centroid $\sigma_j := \frac{\sum_{i=1}^m 1_{\{c^{(i)} = j\}} x^{(i)}}{\sum_{i=1}^m 1_{\{c^{(i)} = j\}}}$
   6. optimizing centroids $J(c^{(1)}, \cdots, c^{(m)}, \sigma_1, \cdots, \sigma_K) = \arg \min_{c^{(1)}, \cdots, c^{(m)}, \sigma_1, \cdots, \sigma_K} \frac{1}{m} \sum_{i=1}^m |x^{(i)} - \sigma_{c^{(i)}}|^2$
   7. $\% \sigma_{c^{(i)}}$ is the cluster centroid to which feature $x^{(i)}$ has been assigned to at last time

After codebook generation, in order to help the security operator to manage which information is contained in the surveilled image data, security operator is enabled to view the patterns in each cluster and assign risk levels to clusters correspondingly. In order to show the result of clustering and risk level assignment, images from UCSD anomaly detection dataset\(^5\) are taken as an example of surveilled data. The UCSD anomaly detection dataset was acquired with a stationary camera, which was monitoring pedestrian walkways. The abnormal behaviors include bikers, skaters, small cars and people walking across a walkway in the grass that surrounds it. After applying pattern retrieval to the image dataset, the extracted patterns are categorized into five clusters, as shown in Figure 24. Figure 24 presents the occurrences of each cluster in one set of the collected dataset. The patterns which are contained in each cluster can be further checked to facilitate risk level assignment for the security operator. For the pattern occurrences shown in Figure 24, the keypoints of the pattern can be displayed, as shown in Figure 25 for the first cluster. Based on the shown keypoints included in each cluster, the security operator can manually assign risk levels to each cluster. Figure 26 assigns the highest risk level to cluster 1 (which clusters most patterns of small cars on the walkway), the high risk level to cluster 2 (which clusters bikers on the walkway and some patterns of small cars), moderate risk level to cluster 5 (which contains most patterns of normal walkers and few patterns from the car), low risk level to cluster 4 and very low risk level to cluster 3 (which contains patterns of the environment of the walkway). After risk level assignment, the occurrence of each cluster will be normalized, as shown in Figure 26.

\(^5\) http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm
6.3.4 Pattern Prediction based on Particle filtering and SIFT Features

Real-time video stream prediction is an important and challenging task. By predicting video streams, the anomalous patterns that are contained in images can be identified and the video surveillance configuration can be optimized by using Hybrid Risk Metrics (HRMs) [50] developed in WP1.

Streaming data is increased when a new frame arrives. This incremental mechanism implies that certain codebook entries will disappear over time (i.e., certain subjects will go out of the coverage of the video camera), while others remains, and new ones emerge. Thus, a forgetting factor is desired to filter historical data. Our approach uses SIFT features of moving subjects as input. All SIFT features (the subject location and its scale) are predicted over time using a particle filter method and their states are updated in each frame. In particle filtering, each state is represented as a particle that has its own state with a structure similar to that of the SIFT feature. The state prediction is realized by using the second order autoregressive dynamics with independent Gaussian noise to compute the new estimated location [51, 52]. By predicting the new location, the next-time step description of the moving subjects is computed around the new location. After predicting SIFT features of next time interval, the improved K-Means method described in Algorithm 1 will again be used to cluster moving patterns and new risk levels will be assigned to each codebook.
6.3.5 Application of Game-theoretical approach (HRMs) to support Video Surveillance

The output of the processed video surveilled data is a probability-distribution (see Figure 26) that can be taken as the payoff of two controllers, the intruder (the subjects, no matter cars, bikers, etc, who is not allowed or restricted to enter a certain area) and the defender (the operator of video surveillance system). Assuming there are two different actions for the intruder to traverse the pedestrian walkway: from south to north and from north to south; and there are also two different actions for the surveillance system to switch on and off cameras: randomly switch on and off and switch on camera for 2 minutes in every 2 minutes. Assuming the 7 days of one week is a repetition for surveillance system to capture video footages (which is a sequence of images). The payoff matrix with raw data of the UCSD anomaly detection dataset is shown in Table 2. In Table 2, each column (corresponding row labeled with 5 -1 risk levels) represents categorical fractions of risk levels (i.e., probability for the clusters). Each column (corresponding row labeled with 5 -1 risk levels) is a distribution over different risk levels and the distribution has the same shape as the curve shown in Figure 26. For each risk level and each day in one week, there is one detection rate. Therefore, for each risk level, there is one distribution of detection rates over the whole week. The detection rate structure remains the same each week. In order to play the game, one naïve approach to aggregate those distributions for one week into a single distribution is to average those distributions with equal probabilities. Therefore, the payoff matrix can be further transform into loss distributions, as shown in Figure 27. The loss distribution shown in Figure 27 is a function of risk levels (from 1 to 5).
Table 2 Payoff matrix of the sample surveillance system

<table>
<thead>
<tr>
<th></th>
<th>From S. to N.</th>
<th>From N. to S.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st</td>
<td>2nd</td>
</tr>
<tr>
<td>Random On&amp;Off</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Every 2 Mins On&amp;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Off for 2 Mins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detec. Rate of risk</td>
<td>0.34</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>0.053</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.217</td>
</tr>
<tr>
<td></td>
<td>0.439</td>
<td>0.196</td>
</tr>
<tr>
<td></td>
<td>0.264</td>
<td>0.244</td>
</tr>
<tr>
<td></td>
<td>0.027</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>0.101</td>
<td>0.291</td>
</tr>
<tr>
<td></td>
<td>0.169</td>
<td>0.244</td>
</tr>
</tbody>
</table>

Figure 27 Loss distribution of the sample surveillance system

After compiling loss distributions, HRMs developed in WP1 can be directly used to give recommendations to security operators in terms of making correct optimal proactive defense actions. D1.2 [52] of the HyRiM project gives out details of the HRMs. Interesting readers can refer to D1.2 for more details about this mathematical framework and we are not going to repeat it here again. The result of applying HRMs on the loss distributions shown in Figure 27 denotes that, the optimal strategies for the surveillance system are to randomly switch on and off the camera with probability one; and the worst case attack strategies for the goal of detection rate is to traverse the pedestrian walkway from north to south with a probability of one.
6.4 Optimal Surveillance Schedule Based on Mobile ID Check technology

In this section, we are addressing surveillance based on the Mobile Id Check Technology. As aforementioned, this physical surveillance needs to mobilize a security guard, and assign to him the task to manually check ids of persons (either employees or visitors...) located in a given area (refer to section 4). Our goal here is to detect potential intruders (persons not known to the system), or “Malicious” employees (known to the system, but with suspicious behavior). The latter represent an internal threat to the utility facility. In our study, we aim at detecting these malicious employees only when they are on the move; they are in a restricted area where they are not allowed to be.

Obviously, the more frequently this manual Id Check is performed, the more efficient we are. However, this comes at a price: the facilities to be monitored are usually spread over relatively large geographic areas. Thus, physical id check through mobile devices held by security guards requires both time (checking a target area may take more or less time depending on its location) and physical effort (we may need more than one single security guard). Besides, frequent id checking may have an impact on the privacy, since employees may feel uncomfortable or always watched. Besides this trade-off, we may also need to find out the optimal strategy (i.e. Schedule) to be applied to meet the goals of this “Efficiency-Cost” balance. This will be tackled through simulation means, presented and described hereafter.

6.4.1 Simulation

As a first step, let us describe our simulation model. We choose to use the INET 3.4 framework [53], on top of OMNeT++ 5.0 discrete event simulator [54] to integrate our model. Through this model, we need to be able to reproduce a faithful image of the physical environment of our monitored facility. We also have to reflect all the applied policies (zone restrictions, employees’ profiles, id check policies, etc.) as well as actors’ behaviors (security guard, simple employee, malicious employee or intruder).

- **The Physical Environment:**

  Our environment consists in a geographic surface, divided into several zones or areas as described in Figure 28. In Figure 28, we can observe several zones (like the one framed in red), reachable through a web of ways/paths to follow when moving from/towards any of these areas. These areas represent the smallest level of granularity of our site. Each of which has an attribute, called “security level”, indicating the criticality of the respective area (a real value ≥ 1).

  In our simulation model, we need to be able to describe our site as a set of areas interconnected through paths. With no lack of genericity, we model each area as a convex polygon, 𝒈, whose center of gravity is located at position \( p = t(x \ y \ z) \), and with orientation (Euler angles [55]) \( \theta = (\alpha \ \beta \ \delta) \) relatively to a fixed reference frame. Note here, if in the real field, our area has a nonconvex shape, it could be divided into two or more neighboring convex shapes (having exactly the same intrinsic characteristics, like the security level), just like depicted in Figure 28 through the blue shaped area. Every area is identified through a unique id, jointly (the area and its id) stored as an entry in a std::map structure, to help accessing it easily. All these information describing our set of areas are presented in an XML file, parsed on the run time, to build and render the physical structure of our site.

  In the same way, paths are modeled as a non-oriented graph \( G = (V, E) \), where \( V \) is the set of vertices, and \( E \) is the set of edges; just as depicted in Figure 29. Vertices in \( V \) represent waypoints, characterized by their geographic coordinates, and corresponding to particular locations in our site, such intersections, area gates, etc. For every couple of vertices \( (e_i, e_j) \in V \times V \), an edge \( (e_i, e_j) \) is added to \( E \) if the two waypoints corresponding to our vertices are directly related by a path in the actual map. It is worth mentioning we are assuming that we can only move straight from waypoint \( e_i \) to waypoint \( e_j \) if they form an edge \( (e_i, e_j) \) in \( E \).
Deliverable 4.3 HyRiM How to Enhance Perimeter Security Using New Surveillance Technologies

Figure 28 Site map sample

Figure 29 Modeling of paths
Thus, the more waypoints we create, the more precise we’ll be. The advantage of such representation, is that we can define one or more weight functions to help select the best (i.e. optimal or near optimal) way to go from one source point (e.g. the head quarter of a security guard) to another destination point (e.g. the gate of a selected area). By default, we are using the hop count as a weight function, but this could be easily extended to real distances between waypoints, or any other weight measure. Note here that we associated to each area a gate, represented as waypoint (i.e. the blue vertex in Figure 29 in V). This could easily be extended to several gates to access the same area (Briefly: add the center of the area as a waypoint, related to all its gates, with the exact same cost (same distance for example). According to the optimality principle of Dijkstra, the optimal way to reach the center of this area will be through the optimal gate to reach this area.). Once again, our graph (i.e. V and E sets) are described in an XML file, parsed on the run time, to build and render the physical structure of those paths in our site. Inside an area, we can have several objects/obstacles, to be avoided during movement. For the time being, this feature is still not implemented in our simulation model.

Figure 30 A built physical environment sample
An example of a built physical structure of a random site is presented in Figure 30. We can see that we defined 3 areas. In our model, we can either give the position of the center of our area and the positions of the outer edge extremities relatively to that center point; or use directly the reference frame to locate those points, in which case, the center is computed automatically on run time. In the example presented in Figure 30, we rather used the second option. We can also notice that besides the positions, security level and gate information of an area, we can specify several additional parameters (color, line type, existence of a fence, etc.) that help customize the final rendering of our environment.

- **Actors:**

In our case study, we can identify two main actor categories: Employees and Intruders. An employee could be either a worker or a security guard. They all hold an ID card meaning that they are known to the system. Unlike an employee, an intruder is someone from outside the facility. So, he either doesn’t hold an ID card, has a fake one, or has a stolen card that does not correspond to his biometrics (i.e. fingerprint or facial photo, etc.). In all these cases, he will not be recognized by the system as a regular employee. Thus, he should be caught at the first ID check, whenever it is done, and wherever he is located inside the facility.

Employees, depending on the job they are supposed to do, are allowed to access some areas of the facility but denied access to some others. This restriction is not always the same for all employees. In our simulation model, we define a set of profiles, each of which indicates a subset of allowed areas. Using an XML file, we assign to each worker one of these profiles, indicating areas he can access. This information is stored in his ID card. Security guards are allowed to access all the areas in the facility. A special profile is then created just for them.

Figure 31 Actors

A regular worker is a person who does respect areas’ restrictions. He will never access an area not figuring in his profile. Thus, upon a security check, his situation would always be fine. In the other hand, a “Malicious” worker is an employee with a valid ID card, but who intends to physically harm the facility. In our work, we
are supposing that such a suspicious behavior consist of targeting areas, probably with high security level, that he is denied access to. During a security check, a malicious worker can only be caught if his is behaving suspicious at that time (i.e. his is in a restricted area when the check takes place).

A security guard owns two main devices: A navigator, and an ID Checker. The navigator serves as a mission scheduler. Checking missions are assigned to a security guard using this device. It first indicates which area a security guard needs to check, shows the way to follow to reach this area, and decides the strategy to be adopted during the ID check. The ID checker is used to verify the identity of an employee. It starts with verifying the ID and the biometrics of the employee. If they match, it verifies whether this employee is allowed to be in the area where the check is performed.

Figure 31 summarizes the hierarchy of actors involved in our simulation. It shows that all of them are able to move in the facility (i.e. they all have a mobility module). Unlike intruders, all employees hold an ID card. Security guards are also equipped with an ID checker and a Navigator devices (they are virtually two separate devices, but could also be integrate into one single physical device). Finally, workers could be of two kinds: regular or malicious.

In our simulation model, a mission consists in three phases:

1. First phase of a mission:
The first phase corresponds to the selection of a target area and guiding the security guard towards it. This selection is made according to a given strategy. At this time, we are basing our choice on the security level of areas in our monitored site. As already said, each area is associated to a security level ($\geq 1$) indicating its criticality (the higher the level, the higher the criticality). Areas are grouped into sets such that two areas, $\mathcal{A}_1$ and $\mathcal{A}_2$ belong to the same set, $\mathcal{S}_i$, if they have the same security level $\ell_i$. As such, all areas in the facility, will be distributed over $m$ disjoint sets, $\mathcal{S}_i$, $i = 1 \ldots m$ verifying $\sum_{i=1}^{m} |\mathcal{S}_i| = n$; where $n$ is the number of areas, and $|S|$ denotes the cardinality of set $S$ (we can easily see that $1 \leq m \leq n$). We can safely assume here that all the sets $\mathcal{S}_i$ are arranged in ascending order; that is $\forall \mathcal{S}_i and \mathcal{S}_j / (i < j) \Rightarrow (\ell_i < \ell_j)$. We inductively define the “range of the security level $\ell_i$” by the equation:

$$range(i) = [\text{lower}_i .. \text{upper}_i] = \begin{cases} \text{upper}_{i-1} .. \text{upper}_{i-1} + \ell_i = \sum_{j=1}^{i} \ell_j & \text{if } i > 1 ; \forall i = 1..m \\
\text{lower}_1 = 0 .. \text{upper}_1 = \ell_1 & \text{if } i = 1 \end{cases}$$

These ranges divide the interval $[0 .. \text{MaxRange}]$, where $\text{MaxRange} = \sum_{i=1}^{m} \ell_i$ (i.e. the sum of all defined security levels in a given scenario) as shown in Figure 32.

![Current strategy to select an area for Check](image)

Based on the obvious rule that a higher security level area should be checked more often, we uniformly generate a value, called $\chi$, in the range $[0 .. \text{MaxRange}]$. Depending on the value of $\chi$, we can decide which set of areas, $\mathcal{S}_i$, we should target now: we go for set $\mathcal{S}_i$ if $\chi \in range(i)$. Thus, a set $\mathcal{S}_i$ is selected with a probability $P_i$ given by:

$$P_i = \frac{\text{upper}_i - \text{lower}_i}{\text{MaxRange}} = \frac{\ell_i}{\sum_{i=1}^{m} \ell_i}$$
Once a set is selected, we uniformly choose to visit one of its areas (if \( \mathbb{G}_i \) is fixed, an area \( A \in \mathbb{G}_i \) is selected with a probability \( 1/|\mathbb{G}_i| \)). This very strategy will be called “Higher Security Levels First” and denoted “HSLF”.

Of course, this is absolutely not the only strategy to be implemented and then evaluated. Other strategies are to be added too. At this stage, we will limit ourselves to simply listing some of them:

- A random selection of areas without paying any attention to the security levels.
- A time-based schedule: at what time each area should be visited?
- A frequency-based schedule: how often an area should be visited?
- Etc.

The navigation device, storing the map of the whole site (i.e. areas and paths), guides the security guard towards the gate of the targeted area. This is done by applying any shortest path algorithm on the graph representing the paths of our site, between the current position (the head quarter for the security guard) and the gate of the area to be checked.

- Second phase of a mission:

The second phase of a mission is checking the selected area. The security guard needs to move all around and meet workers located in this area for an eventual ID check. Inside an area, we can apply any of the mobility models provided by INET framework. Yet, we may need to extend them since it only support movement inside a squared area. This work is already done for one of these mobility models, which is the well-known Random Waypoint mobility model [56, 57]. Basically, a mobile node inside a selected area, uniformly generates a target position inside the polygon surface of our area, selects a speed, and then moves towards its target. At its arrival, the node waits at its position for a randomly generated time, before reproducing the same behavior once again. Notice here that all our actors are moving with respect to this same mobility model. The only difference might be the “move-wait” pattern. In fact, a worker would spend most of its time in the same place doing his work, then moves to another place to do some other work, and so on and so forth. On the other hand, a security guard would spend most of his time moving from one position to another, with very short and brief waits. A malicious guy, either an intruder or a worker, would be moving more than a regular worker, but spending more time waiting as he is supposed to do some harmful work. All these parameters could be fixed as input to our scenario.

While moving, a security guard will meet persons who are in the checked area. For every one in his direct vicinity (i.e. within a radius \( R \) of \( r \) meters), the security guard decides to check his ID with a given probability (by default, the probability is set to 0.5). This probability should be closely related to the security level of the area. Every selected subject, remains at his current place until the check is performed. If a malicious person (i.e. intruder or worker) is detected, a “handle situation” procedure is triggered. This procedure could be of any type, like calling a third party to drive the caught guy to an interrogation room, or the security guard stopping the checking mission and driving the checked person back to the head quarter by himself, or since we are running a simulation, remove the malicious node from the simulation and continue the checking mission, or just simply stopping the simulation.

Besides, to avoid that a security guard checks repeatedly the same person again and again during one same check spot mission, we added a memory module to the security guard. This module, being tunable, will decide the behavior of a security guard according three basic features: how easily can he remember a new face? For how much time can he keep remembering it? And, how many faces can he remember? The first feature, called “the memory quality”, is a probability like parameter to be given as an input: it lays between 0, meaning that he can’t remember anything, and 1, meaning he remembers everything. The second feature, called “the memory time”, is a time duration to be given as an input. It can either be a fixed duration or a distribution (e.g. a uniform distribution) which indicates for how much time a newly met face could be remembered. Every new entry to the memory will be associated to “memory time” value to decide when to
forget it. The third feature, represents the size of the memory and hence called “the memory size”. It is implemented as a circular buffer, so that if it is full, the oldest face (having the smallest “memory time” value) would be forgotten first.

The end of the second phase depends on the applied strategy. It can end after a time duration spent inside the area, after a fixed number of checked persons is reached, or after the checking ratio goes beyond a given threshold (if the number of workers inside the area is a priori known). In any of these cases, the security guard announces the end of this phase using his navigator device. And the mission shifts to its third and final phase.

Third phase of a mission:

This phase is the easiest and simplest phase to describe. It only consists on guiding the security guard back to the headquarter using the reserve path stored in the navigation device. The security guard needs to empty his memory, because he should be able, in the upcoming missions, to re-check a person already checked, also that person could move from an area to another at any time.

6.4.2 Scenario

6.4.2.1 Scenario description

![Figure 33 screenshot of the simulated area](image)

The simulated physical environment, as illustrated in Figure 33, consists of an area (500 m x 250m), divided into twelve sub-areas denoted A1, A2 … A12 respectively. To each sub-area, we assign a security level indicating how critical the sub-area is: the higher the security level, the more critical the sub-area is. The respective assigned security levels are described in Table 3

<table>
<thead>
<tr>
<th>Table 3 Security levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
</tr>
<tr>
<td>Security Level</td>
</tr>
</tbody>
</table>
These levels are used to evaluate the caused damage of an attack. It is obvious that the caused damage would be proportional to the time spent in a sub-area and its security level. This relation is expressed in the following equation:

\[
causedDamage = \frac{1}{NI} \sum_{i=1}^{NI} \sum_{j=1}^{12} timeSpent(i, A_j) \times secLevel(A_j)
\]

where \(NI\) is the total number of attackers, \(timeSpent(i, A_j)\) is the time spent by attacker \(i\) inside sub-area \(A_j\) before being caught by a security guard, and \(secLevel(A_j)\) is the security level assigned to sub-area \(A_j\).

Recall here that we always intend to minimize the caused damage of any given attack. This can be achieved by detecting intruders (attackers) quickly, and thus increasing the number of security guards, or focusing checking missions on sub-areas with higher risk (recall the "HSLF" strategy). However, this could indeed increase the detection rate (given by the next equation) or reduce the caused damage at the cost of affecting the comfort of the facility employees as they would be checked several times a day and thus feel like working under continuous control (could be seen as a privacy issue). Therefore, an optimal multi goal strategy should rather be selected to win such a trade-off.

\[
DetectionRate = \frac{\text{number of detected attackers}}{\text{Total number of attackers}}
\]

Of course, the more a worker is checked, the more uncomfortable he’ll be. However, it is still a subjective feeling with regards to when starting to feel and how much uncomfortable an employee would be after several checks. Consequently, and to measure the privacy/comfort breach of the employees, we made an empirical study through a questionnaire distributed among Akhela employees of a critical infrastructure facility, where we asked about their feeling (scored between 0 and 1, where 0 is a total comfort preservation and 1 means a maximum comfort breach) if ever they get checked 1, 2, ..., 9 times (or more) a day. The collected data showed us a general intuitive curve as shown in Figure 34.

![Figure 34 the comfort breach curve](image)

Using this data, we created a multivariate Gaussian \((\mu, \Sigma)\) [58, 59] used in our simulator as a simple generator so that we can create as many workers as we want, with different subjective comfort breach measures, but following the same general shape as the one shown in Figure 34.

Obviously, we can define several additional key performance indicators (KPIs) and respective target goals such as resource cost, energy cost of the mobile checking devices, etc. However, we will limit ourselves to the three aforementioned measures: (i) detection rate, (ii) employees’ comfort breach, and (iii) caused damage. Our ultimate goal is to show the multi-goal aspect to be considered while defining the optimal strategy.
Every employee holds an ID-card proving his identity and his right to be in a given sub-area. They can move between sub-areas following the layout’s paths and ways connecting them. They are also free to move inside a sub-area according to the following movement pattern: (i) select a random position inside the sub-area, (ii) move to this position (iii) spend some time (a stay) working in this position. This stay is chosen to be uniformly distributed between 10 and 60 minutes. As aforementioned, every security guard follows a schedule of checking missions where he is supposed to move around and check the identity of some randomly selected employees in the zone. A schedule indicates when a mission should start. The number of checking missions is equal to $N_{missions}$ per security guard per day, uniformly spread over the 8 working hours. Every mission lasts for a duration between $MinDuration$ and $MaxDuration$ minutes, during which a security guard may check the identity of several employees that he meets (i.e. in his direct vicinity set to a radius of 5 meters) in his way. A checking operation may last between 20 seconds and 1 minute.

On the other hand, intruders are not authorized to be in any of the zones of our company. They are holding fake ID-cards, and are certainly detected upon a check operation. An intruder may choose to remain in the sub-area where he is, or move from one sub-area to another following a given strategy (i.e. randomly or “HSLF”). At the cost of being possibly detected by a security guard, staying in the same zone means adopting a movement pattern similar to a regular employee but with shorter stay time since he has no actual work to do (we set this value to 30s). The most important parameters of our simulation are summarized in Table 4:

<table>
<thead>
<tr>
<th>Table 4 simulation set up (General parameters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation Time</td>
</tr>
<tr>
<td>Nb of areas</td>
</tr>
<tr>
<td>Nb of employees</td>
</tr>
<tr>
<td>Check time</td>
</tr>
</tbody>
</table>

For the defender side, as well as for the attackers’ side, several strategies were simulated. These strategies are summarized in the following Table 5 and Table 6, respectively:

<table>
<thead>
<tr>
<th>Table 5 Defender strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Defender Strategies</strong></td>
</tr>
<tr>
<td>Number of security Guards</td>
</tr>
<tr>
<td>Missions duration</td>
</tr>
<tr>
<td>Missions frequency</td>
</tr>
<tr>
<td>How to select which area to visit now?</td>
</tr>
<tr>
<td>1. Choose randomly</td>
</tr>
<tr>
<td>2. The higher the security level, the most probably the area is selected</td>
</tr>
<tr>
<td>3. Fixed sequence</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6 Attacker strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attacker Strategies</strong></td>
</tr>
<tr>
<td>Number of intruders</td>
</tr>
<tr>
<td>How to select which area to target now?</td>
</tr>
<tr>
<td>1. Choose randomly</td>
</tr>
<tr>
<td>2. The higher the security level, the most probably the area is selected</td>
</tr>
<tr>
<td>3. The lower the security level, the most probably the area is selected</td>
</tr>
</tbody>
</table>

Some of the collected simulation results are presented in the following section.

**6.4.2.2 Simulation results**

For the sake of simplicity, we choose to consider only the following strategies in Table 7:
As explained earlier, for our scenario, the strategies are based on the number of security guards, the frequency at which a security guard will visit the zones, and which areas are targeted first in one hand; and the number of attackers and which areas will be targeted first by an attacker in the other hand. For the simulation we are using an abstract notation encoding these parameters on both sides, i.e., for the defender and the attacker. Defensive strategies are marked with a prefix “D” and are of the form “D- NGiFFTc”. The letters indicate the meaning of the numbers or descriptors, i.e.:

- \( NGi \): Number of security guards is \( i \).
- \( FFf \): Frequency of checks is \( f \). That means that the number of checking missions is \( f \) uniformly distributed over the whole working day, i.e. 8 working hours.
- \( Tc \): How the area is targeted, i.e., either randomly \( (c = R) \), or higher security levels more frequently \( (c = HSLF) \).

Likewise, attack strategies are described by strings prefixed with “A” and have the general form “A- NiITc”.

Similar to the defensive strategies, this encoding has the following meaning:

- \( Ni \): Number of intruders is \( i \).
- \( Tc \): How the area is targeted, i.e., either randomly \( (c = R) \) or higher security levels more frequently \( (c = HSLF) \).

For our simulations, we used three different values for the number of security guards (1, 3 and 5) as well as the number of intruders (1, 5 and 10) and used two different frequencies for the security guards (3 and 7). For all possible combinations, we evaluated both types of targeting strategies for the defenders (i.e., targeting zones randomly or high security zones more frequently) and only the random strategy for the attacker. This results in six defensive strategies and three attack strategies, as given in Table 7 and Table 8.

In our test scenario, not only one single goal is of interest but we have to take several aspects into consideration to find an optimal solution for the game. In more details, we have limited ourselves to three measures only: the maximum privacy/comfort breach, the caused damage, and the detection rate. The overall game has three goals: Damage, Privacy and Detection Rate. These three goals have to be minimized (e.g., Damage) and maximized (e.g., Detection Rate) at the same time based on the different attack and defense strategies (cf. Table 7). As already described in section 6.2, our game-theoretic framework is able to solve such a multi-objective game.

Table 7 Considered strategies

| 3 defender strategies (Missions duration: uniform(5min, 15min)) | 1 sec. guard, freq = 7 & areas: targeted randomly
| D-NG1F7TR | |
| D-NG3F7TR | 3 sec. guards, freq = 7 & areas: targeted randomly
| D-NG5F7TR | 5 sec. guards, freq = 7 & areas: targeted randomly
| D-NG1F7HSLF | 1 sec. guard, freq = 7 & areas: targeted Higher Sec. Lev. First
| D-NG3F7HSLF | 3 sec. guards, freq = 7 & areas: targeted Higher Sec. Lev. First
| D-NG5F7HSLF | 5 sec. guards, freq = 7 & areas: targeted Higher Sec. Lev. First

| 3 attacker strategies | 1 intruder & areas are targeted randomly
| A-NI1TR | |
| A-NI5TR | 5 intruders & areas are targeted randomly
| A-NI10TR | 10 intruders & areas are targeted randomly

As already described in section 6.2, our game-theoretic framework is able to solve such a multi-objective game.
In a next step, the damage caused by the attacker has to be evaluated for all three of the aforementioned goals. This has to be done for every combination of attack and defense strategy and the individual results can be obtained, e.g., by using a certain number of simulations. In a conventional approach, the results of these simulation would be averaged to get a single value for each of the three goals. Further, these averaged values would be used to solve the so-arising multi-objective game using standard methods. In our approach, we deviate from this route by not aggregating the data and calculating the average, but instead compiling all output of the simulation in categorical distributions (effectively histograms), over which the game payoff structures (one per goal) are defined. At the current state, we only aim at explaining how results are interpreted and understood rather than having an extended set of simulations. Recall here also that every scenario was run for 5 times. The results will be then categorized into 5 fixed classes as presented in Table 8:

<table>
<thead>
<tr>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

very low low medium high very high

This categorization is mandatory to be able to apply the HyRiM game-theoretical framework capable of computing the optimal strategy to be applied out of our multi-goal game [50]: This is simply done by diving the resulting data into these five categories that equidistantly span the numeric range of all three goals, forming a histogram for each result. The respective matrices displaying all this information, i.e., all the histograms (payoff structures for the various goals), are plotted in ANNEX II: Payoff Structures for goals damage, detect rate, and privacy of the paper as Figure 47, Figure 48 and Figure 49. The raw results are also given by Table 9:

<table>
<thead>
<tr>
<th>Table 9 D-NGXFTR vs. A-NIXTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-NI1TR</td>
</tr>
<tr>
<td>Run #</td>
</tr>
<tr>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>Max. Privacy</td>
</tr>
<tr>
<td>D. NGXFTTR</td>
</tr>
<tr>
<td>NPS</td>
</tr>
</tbody>
</table>

As we can see through Table 10, as we increase the number of security guards, the detection rate increases but at the cost of increasing the privacy breach as explained in the previous section. This proves that choosing an optimal strategy is not an easy task since it has to take into account several tradeoffs. Through Table 10, we can also see that when the defender chooses to apply a “HSLF” strategy rather an random strategy, he may achieve better results in terms of caused damage (the damage is reduced), but at the same time, the maximum privacy breach increases as security guards will target more often the same areas where the same employees will be checked repeatedly during the same day.
### 6.4.2.3 Application of Game-Theoretical Framework (HRMs) to support the Mobile-ID-Check technology

As we can see in Table 9 and Table 10, the output of our simulation is sets of data that can be seen as a distribution for each couple of (player1: defender, player2: attacker) strategies. These distributions are in fact the payoff of our two players. Therefore, HRMs developed in WP1 can be directly used to give recommendations regarding the best strategy, or more correctly the best mixed strategy, to be applied from the defender point of view, as well as the potential damage that can be caused by a worst case attack. Now, we can apply our game to decide on what is optimal for us in this very scenario. (For more details about this HRMs mathematical framework, please refer to D1.2 [50] of HyRiM project), where a fictitious play algorithm was used that iterated until a precision of 0.001 was reached to compute the equilibrium distribution approximation in the t-th iteration. As said, in general, the resulting solution is a mixture of all possible security strategies specified in the first step (cf., Table 7). In other words, each strategy of the solution has a specific probability to be carried out. Computing the equilibrium, the optimal security strategies for both players are illustrated in Figure 35 and Figure 36.

After categorizing data, the loss distributions over all the goals have been compiled. We considered even priorities among all the different goals. As expected, the results delivered a nontrivial mixed strategy. This means that almost all defense strategies are useful, except for strategy #4 "D-NG1F7THSLF", which has probability equal to zero assigned to it as illustrated in Figure 35. Hence, a practitioner could abandon such a strategy to be as an option to defend - as such, this is also an interesting observation/lesson from the game: it also tells us which defenses are more relevant compared to others. The equilibrium shows the most important/effective countermeasures (being “D-NG3F7TR”, followed by “D-NG5F7THSLF”, which are sufficient >90% of the time, until “D-NG1F7THSLF”, which turned out to be completely useless). To summarize, for the present scenario, our best option is a mixture of five defense strategies, i.e., strategies #1, #2, #3, #5 and #6 (cf. Table 7). In more detail, strategies #1 (“D-NG1F7TR”), #2 (“D-NG3F7TR”), #3 (“D-NG5F7TR”), #5 (“D-NG3F7THSLF”) and #6 (“D-NG5F7THSLF”) have to be applied with respective probabilities 0.001, 0.809, 0.074, 0.006 and 0.11.
In addition, the worst-case attack strategies are different between the three goals, as shown in Figure 36, so the attacker can never cause maximal loss in all three goals at the same time. He can only cause a maximum loss in one single goal a time. Such an observation represents good news, since it indicates that the computed loss distributions are only pessimistic and reality should look much better (expectedly).

As already stated above, the method for computing this equilibrium is essentially based on multicriteria optimization for security (cf. section 6.2). Provided that the defender plays this optimal strategy, the respective optimal loss distributions attained under any behavior of the attacker are given in Figure 37, Figure 38 and Figure 39.
Deliverable 4.3 HyRiM How to Enhance Perimeter Security Using New Surveillance Technologies

Figure 37 Assurance Plots of Damage

Figure 38 Assurance Plots of Detect Rate
Finally, we may add here that the results from the game-theoretic optimization algorithm need to be implemented precisely since a deviation of the probabilities given in the equilibrium will increase the potential damage caused by an attacker under the worst case attack strategy. Therefore, the results from the algorithm can be fed into a random selection function to obtain the current advice according to the optimal randomized choice rule for the strategies. In other words, the manager of the security guards will use some kind of scheduling system, which provides him with an indication which of the five security strategies (either #1, #2, #3, #5 or #6) to follow at each time. This needs to be done iteratively (e.g., each day or at the beginning of each shift) and can be precomputed for several days or weeks to simplify personnel decisions (e.g., shift rotations, etc.).

6.5 A Privacy Breach Evaluation Model on Mobile ID Check Technology

Nowadays, surveillance systems play a very important role in critical infrastructures where security and safety are vital factors. However, and despite the very strict and rigid applied standards to stop and prevent such attacks, especially applied at the outer perimeter of the protected facility, some attackers still succeed in bypassing this outer security control and thus are able to cause tremendous damage. In most cases, such situations are not tolerated especially when it is about a critical infrastructure. Thus, control and security measures are extended to the inside of the facility. For that, more flexible and resilient surveillance methods and operations are adopted, e.g. on-demand surveillance. In fact, to be able to apply such an on-demand surveillance, new non-traditional devices are used, which are usually portable and wireless for sake of mobility capability. Such devices can be operated when and where it is deemed necessary. The whole applied system can be briefly described as a server, one or more on-demand devices acting as trusted computing platforms [35] and objects to be identified. As a first operational phase, we analyze possible risks and set accordingly surveilling strategies. The hand-held devices are used following these fixed strategies to perform checks on different objects or subjects (employees for instance), then report the surveilling data to the server. Finally, the server analyzes this data and finds latent risks.
6.5.1 Overview of Location Privacy

As explained, on-demand surveillance devices help detect threats that succeeded in bypassing the control traditionally located at the border. Thus, intruders can be detected even if they managed to enter the facility without being spotted by the applied security system at the entrance. For that, all subjects inside the facility can, at any time, be asked to verify their identity: “is subject \( i \) allowed to be in this position?”. Unfortunately, such a procedure has a major drawback: revealing locations of authorized subjects, breaching consequently their location privacy. Location privacy is a particular type of information privacy that we define as the ability to learn all subjects’ locations. Such information represents a threat itself if revealed. In fact, attackers knowing about this revealed subjects’ locations at a particular time can, for instance, infer their identity using powerful data mining approaches, or plan an attack based on subjects movements after inferring it from these subjects’ locations.

In our daily life, examples of location privacy violation are numerous, including social networking with geo-tagged messages, video surveillance, local business search, automotive traffic monitoring, ubiquitous surveillance (e.g. smartphones, wearable digital media and ICT implants) and so on. Those privacy violations can roughly be categorized as snapshots and continuous location privacy, which are depicted in Table 11.

<table>
<thead>
<tr>
<th>Privacy Violation</th>
<th>Technologies</th>
<th>Privacy Threats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snapshot</td>
<td>Social Networking</td>
<td>Shared messages are with geo-information</td>
</tr>
<tr>
<td></td>
<td>Video Surveillance</td>
<td>Visually reporting geo-information</td>
</tr>
<tr>
<td></td>
<td>Local Business Search</td>
<td>Searching for stores within a range distance</td>
</tr>
<tr>
<td>Continuous</td>
<td>Automotive Traffic Monitoring</td>
<td>Inferring traffic congestions from location</td>
</tr>
<tr>
<td></td>
<td>Ubiquitous Services</td>
<td>Continuously reporting geo-information</td>
</tr>
</tbody>
</table>

For the snapshot privacy violation, a mobile subject only needs to report his current location to a service provider once to get his desired information. And for continuous privacy violation, the subject has to report his location to a service provider in a periodic or on-demand manner. Both snapshot and continuous location information can invade individuals’ privacy seriously. In fact, according to Marist Poll [60] and Webroot [61], 50% of U.S. residents who have a problem on a social networking site are concerned about their privacy and more than 55% of location-based service users show concern about their privacy violation. Many real-life scenarios where perpetrators abuse location-detection technologies to gain access to private location information about victims, e.g. the employers could, from analyzing the individual data of the participant subjects determine that they have not been working as one would expect, took more time to do the rounds, that they have spent more time in the lavatory (for smoking for example) than permitted, and thus take disciplinary actions.

To conclude, this location privacy is also a problem we are facing using this mobile Id check technology, as, and even if the aim of this applied surveillance is only to identify subjects, it exposes subjects’ locations as well. As a result, it is a must that we focus on the privacy breach evaluation on using location information in on-demand surveillances. When on-demand surveillance devices are used, an enormous amount of potentially sensitive information is generated. We do not necessarily want to rule out any privacy breach, but evaluate how much privacy information still remains in the systems.
6.5.2 Previous Evaluation Approaches

Besides privacy preservation technologies, privacy evaluation technologies are another important privacy study branches. The goal of privacy metrics is to measure the privacy impact in a system. Location privacy was and still is a popular topic in the last decade due to its importance. Current privacy evaluation metrics can be categorized as: expectation, time, accuracy and uncertain, based on the technologies that the privacy metrics employ.

An intuitive approach on privacy breach evaluation is to calculate the expectation of privacy as determined by attackers. However, this metric does not indicate the severity of a leak. k-anonymity model [62] opens a novel way to preserve privacy over data publishing. It also proposes an approach to calculate expected similarity among the published items. After that, l-diversity [63] and t-closeness [64] were proposed to improve k-anonymity, but they share the same basic principles. The evaluation approach is simple and general. This model only depends on the number of elements in the model, and does not take into account the information that attackers may obtain by observing the system for a while.

The time needed by attackers to be spent to compromise subjects’ privacy is also an important metric to evaluate the privacy of a system. Those metrics measure the time until attackers succeed, including maximum tracking time [65] and mean tracking time [66]. Those approaches are mainly proposed for anonymous communication. After multiple rounds, attackers can successfully get sources of data due to some fixed information [67].

On the other hand, accuracy metrics quantify the accuracy of attackers’ estimation. Confidence interval width [68], mean squared error [69], distance error [70] and normalized variance [71] are important approaches employed to evaluate the accuracy of attackers’ location estimation. The amount of privacy at a percentage is given by the width of the interval between the estimation and the true outcome. However, those approaches require subjects’ location to get predefined metrics.

In general, Entropy is a measure of unpredictability of information content [72] [73]. If the content is certain, the entropy is zero and the outcome can be predicted perfectly. As entropy increases, the prediction becomes more and more imperfect. For this very reason, entropy theory is one of the most popular approaches in privacy evaluations.

6.5.3 Assumptions

To elaborate our model, we assume the following:
1. The monitored area is divided into multiple sub-areas. The sub-areas are not overlapping.
2. Each subject has a unique identity.
3. In each sub-area, random checks using on-demand devices, are performed in a complete independent way from each other’s.
4. Attackers are always aware of the number of subjects inside the whole monitored area.

6.5.4 Notations

Following, we list some of the most important notations used to build our privacy model:
1. $\mathcal{A}$: set of areas composing the facility. The number of areas is denoted by $|\mathcal{A}| = n$.
2. $\mathcal{S}$: set of subjects (either employees or attackers) present in the facility. The number of subjects is denoted by $|\mathcal{S}| = m$.
3. $\mathcal{H}$: set of subjects being checked. $\mathcal{H} \subseteq \mathcal{S}$.
4. $\mathcal{O}$: set of events $\sigma_{i,j}$ $i = 1..m$ $j = 1..n$: subject $i$ is in area $j$. 
6.5.5 Attackers

The metric of privacy breach depend on the probabilities that the attacker can successfully get each subjects' location (in which sub-area is he located). The metric is therefore measured with respect to a particular attack: how much information is still needed to totally get all subjects' privacy. Concrete assumptions about the attacker have to be clearly specified when measuring the metrics of privacy breach.

1. External attackers can only eavesdrop surveilling data from mobile check devices. They can observe subjects getting in and out the monitored area and listen to the communication between devices and servers. They know the total number of subjects and sub-areas, but they do not know subjects initial locations (in which sub-area is every subject located).
2. Internal attackers, e.g. malicious security guards: besides the surveilling data they hold legitimately, these attackers also have other background regarding subjects' original locations and behavior pattern.

6.5.6 Evaluation Model

The aim of the model is to provide the metrics of privacy breach caused by the on-demand surveillance system.

6.5.6.1 Basic Model against External Attackers

In this subsection, we introduce a basic evaluation model supposing external attackers. The subjects’ possible locations can be treated as data sources in the communication model, as described in Figure 40. At the beginning, events “subject $i$ is in a sub-area $j$” are independent. The attacker can eavesdrop communication channel and get some location information exchanged between the server and surveillance devices.

![Communication model of privacy breach with external attackers](image)

His goal is to predict the location of all subjects with a high degree of confidence. In other words, he is attempting to find/guess the probabilities given by the probability matrix presented in equation (4):

$$P(O) = \begin{pmatrix} 1 \\ \vdots \\ m \end{pmatrix} \begin{pmatrix} 1 & \cdots & n \\ p(\sigma_{1,1}) & \cdots & p(\sigma_{1,n}) \\ \vdots & \ddots & \vdots \\ p(\sigma_{m,1}) & \cdots & p(\sigma_{m,n}) \end{pmatrix}$$

(4)

where, $0 \leq p(\sigma_{i,j}) \leq 1 \ \forall \ i = 1..m, \ \forall \ j = 1..n$, and $\sum_{j=1}^{n} p(\sigma_{i,j}) = 1, \ \forall \ i = 1..m$. 

69
For the time being, we will only focus on a **Snapshot** privacy violation case, where we suppose that one single checking mission was performed per area (we insist here that in one mission, we can check more than one subject.), and that we are interested in measuring the possible privacy breach afterwards.

Before going any further, let us start by taking a closer look at the Shannon entropy.

**Definition 1:** *Shannon entropy:* We define the Shannon entropy of $\mathcal{O}$ by:

$$H(\mathcal{O}) = -\sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \cdots \sum_{j_m=1}^{n} \prod_{i=1}^{m} p(\sigma_{i,j_i}) \log_2 \prod_{i=1}^{m} p(\sigma_{i,j_i}) \quad (5)$$

Here, we can legitimately suppose that for any subject $i$, the probability of being in any zone is evenly distributed (Hypothesis H1) [74].

**Hypothesis:**

\[
\forall (j_1, j_2) \in \mathcal{A}^2, \forall i \notin \mathcal{H} \quad p(\sigma_{i,j_1}) = p(\sigma_{i,j_2})
\]

Leading to:

\[
p(\sigma_{i,j}) = \frac{1}{n} \forall i \notin \mathcal{H}, \forall j = 1..n \quad (H1)
\]

To better understand this entropy, we will consider the following two illustrative examples:

**Simple example:**

Let $n = 2$ (2 areas denoted A and B) and $m = 3$ (3 subjects denoted 1, 2 and 3).

1. **If $\mathcal{H} = \emptyset$**

   \[
P(\mathcal{O}) = \frac{1}{2} \begin{pmatrix} A & B \\ 1/2 & 1/2 \\ 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}
   \]

   And then,

   \[
   H(\mathcal{O}) = - \sum_{f_1=1}^2 \sum_{f_2=1}^2 \sum_{f_3=1}^2 \prod_{i=1}^3 p(\sigma_{i,j_i}) \log_2 \prod_{i=1}^3 p(\sigma_{i,j_i})
   \]

70
Let's say that subject 1 was checked in A.

Consequently:

\[
H(O) = \left(- \frac{1}{2} \times \log_2 \left(\frac{1}{2}\right)^3 + \cdots + \frac{1}{2} \times \log_2 \left(\frac{1}{2}\right)^3\right) = \log_2 2^3
\]

Under (H1):

\[
H(O) = -\sum_{k=1}^{m} \left(\frac{1}{n} \times \cdots \times \frac{1}{n}\right) \times \log_2 \left(\frac{1}{n} \times \cdots \times \frac{1}{n}\right) = \log_2 n^m
\]  

(6)

2. If \( k = \{1\} \)

Let's say that subject 1 was checked in A.

We only have 4 combinations:

\[
P(O) = \begin{pmatrix}
\frac{1}{2} & \frac{1}{2} \\
\frac{1}{2} & \frac{1}{2} \\
\end{pmatrix}
\]

Consequently:

\[
H(O) = \left(- \frac{1}{2} \times \log_2 \left(\frac{1}{2}\right)^3 + \cdots + \frac{1}{2} \times \log_2 \left(\frac{1}{2}\right)^3\right) = \log_2 2^2
\]

To generalize:

If \( S \) is the set of all subjects, \( \bar{k} \) is the set of checked subjects and \( \bar{S} = S \setminus \bar{k} \) is the set of non-checked subjects.

Let \( |S| = m \), \( |k| = m' \) and \( |\bar{k}| = m - m' = m'' \).
Deliverable 4.3 HyRiM How to Enhance Perimeter Security Using New Surveillance Technologies

\[ P^*(\mathcal{O}) = \begin{pmatrix} 1 & \cdots & \cdots & \cdots & 1 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ m'' & \cdots & \cdots & \cdots & m'' \end{pmatrix} \]

where \( p^*(\sigma_{i,j}) = p(\sigma_{i',j}) \) \( \forall i \in \tilde{h}, i' \in S \text{ and } i = i' \) (same subject)

\[ H^*(\mathcal{O}) = -\sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \ldots \sum_{j_m=1}^{n} \prod_{i=1}^{m''} p^*(\sigma_{i,j}) \log_2 \prod_{i=1}^{m''} p^*(\sigma_{i,j}) \]  

Under H1:

\[ H^*(\mathcal{O}) = \log_2 n^{m''} = \log_2 n^{m-m'} \]  

**Remark:**

Under H1:

- If \( \tilde{h} = \phi \), \( H^*(\mathcal{O}) = H(\mathcal{O}) \) and \( P^*(\mathcal{O}) = P(\mathcal{O}) \)
- If \( \tilde{h} \neq \phi \), \( H^*(\mathcal{O}) < H(\mathcal{O}) \) and \( P^*(\mathcal{O}) \neq P(\mathcal{O}) \)

In fact, for any given subject \( i \), being in any location is equally likely, which corresponds to the maximum of uncertainty. Any deviation from this probability means less uncertainty and should be then quantified in lower entropy. The extreme case is when all employees are checked, which means there is no uncertainty at all and thus the entropy is equal to zero.

**Definition 2:** Normalized Shannon entropy: Under H1, we define the normalized Shannon entropy of \( \mathcal{O} \) by \([75]\):

\[ V(\mathcal{O}) = \frac{H^*(\mathcal{O})}{H(\mathcal{O})} = \frac{\log_2 n^{m-m'}}{\log_2 n^{m}} = \frac{m-m'}{m} \]  

Again, we can observe that:

- If \( \tilde{h} = \phi \), \( V(\mathcal{O}) = 1 \) : maximum of privacy preservation
- If \( \tilde{h} \neq \phi \), \( V(\mathcal{O}) < 1 \) : privacy is breached
- As \( |\tilde{h}| = m' \\
\), \( V(\mathcal{O}) \to 0 \) : maximum of privacy breach.

Thus, \( V(\mathcal{O}) \) could be a good metric for privacy preservation.

**6.5.6.2 Evaluation Model against Internal Attackers**

Internal attackers have more information regarding subjects. In a continuous privacy violation scenario, such attackers could guess with a better confidence the location of subjects even a while after being checked (better that the mere even guess \( \frac{1}{n} \)). The communication model would be like Figure 41. Attackers can directly get surveilling data from devices, but also rely on their experience to reveal subjects’ locations.
As said, we will try here to consider a continuous privacy violation model. That means we want to measure the privacy breach/preservation as a function of time. In fact, from an attacker point of view, predicting the location of an employee shortly after checking him would be with more confidence compared to a prediction made a long time after the last check. This is due to the fact that the acquired data at a given instant (right after the check) would lose its value over time and slowly converge to the steady (equally likely) situation. As such, the respective probabilities of a subject $i$ to be IN or OUT a given area $j$ respectively should rather evolve over time as illustrated in Figure 42:

![Figure 42 Probability of an Event](image-url)

As time ticks, subject $i$, who was checked in area $j$ at $t_1$, could leave this area and move to another place. Thus, $p(o_{i,j})$ and $p(o_{i,j'}\neq j)$, which were at $t_2$ equal to 1 and 0 respectively, should decrease and increase respectively over time until they reach the probability of $\frac{1}{n}$ (under H1).

In this section, we will investigate the expression of $p(o_{i,j})$, $p(o_{i,j'}\neq j)$ as functions of time, and $\Delta t$, the time after which the probability $\frac{1}{n}$ (maximum uncertainty) is reached.

Recall here that the exponential distribution is usually used to model time between the occurrences of events in a time interval. It is an appropriate model if the following conditions are true:

- X is the time between events, with $X>0$.
- If event e1 occurs, it will not affect the probability of occurrence of a second event e2 (all events are independent).
- The rate at which events occur is constant (do not vary over time).
- 2 events can’t occur at the exact same time.

Let’s for now suppose that we have one single subject $i$. If we suppose that the movement pattern of such a subject from one area to another remains the same over the whole working day (the subject behaves exactly...
the same way for the whole day and his behavior is not affected by special times such as the lunch time), we can safely suppose that the stay duration $T_j^\text{in}$ of subject $i$ in a given area $j$ (time spent between two successive “enter area $j$” and “leave area $j$” events) follows an exponential distribution of parameter $\lambda_{i,j}$ ($T_j^\text{in} \sim \exp(\lambda_{i,j})$). Similarly, we can also suppose that the duration spent outside the area $j$, denoted $T_j^\text{out}$, follows an exponential distribution of parameter $\mu_{i,j}$ ($T_j^\text{out} \sim \exp(\mu_{i,j})$). Otherwise, we should divide the working day over different time intervals such that of each individual time interval, our hypotheses are kept correct.

**Hypotheses:**
1. $\forall i \in S$, and for a given area $j \in \mathcal{A}$:
   - The “in” duration of subject $i$ in area $j$ is the random variable $T_j^\text{in} \sim \exp(\lambda_{i,j})$
   - The “out” duration of subject $i$ outside area $j$ is the random variable $T_j^\text{out} \sim \exp(\mu_{i,j})$ (H2)
2. $|\mathcal{A}| = n > 1$, otherwise $\sigma_{i,1}$ will always be true $\forall i$ and for all time. (H3)

Under hypothesis (H2), we can assimilate each area $j$ as a Markov Chain of 2 states: $I_{i,j}$ and $O_{i,j}$ (see Figure 43).

![Two State Markov Chain Representing Subject Movement From/To a given area](image)

Figure 43 Two State Markov Chain Representing Subject Movement From/To a given area

And thus, we can describe the temporal evolution of our system, $(X_t)_t$, using this two-state C.M.C.

Let’s start by computing the transition matrix and the process generator $A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}$.

For $t$ small:
- If we suppose that subject $i$ was initially outside area $j$:
  $$P(X_t = O_{i,j}|X_0 = O_{i,j}) \approx P(T_j^\text{out} > t) = e^{-\mu_{i,j}t}, \text{ leading to } a_{2,2} = \frac{dP(x_i = O_{i,j}|x_0 = O_{i,j})}{dx_i} \bigg|_{x_i = 0} = -\mu_{i,j}$$
- If we suppose that subject $i$ was initially in area $j$, then:
  $$P(X_t = I_{i,j}|X_0 = I_{i,j}) \approx P(T_j^\text{in} > t) = e^{-\lambda_{i,j}t}, \text{ leading to } a_{1,1} = \frac{dP(x_i = I_{i,j}|x_0 = I_{i,j})}{dx_i} \bigg|_{x_i = 0} = -\lambda_{i,j}$$

For the sake of clarity, we will note $\lambda_{i,j}$ simply by $\lambda$, $\mu_{i,j}$ by $\mu$, $I_{i,j}$ by $I$ and $O_{i,j}$ by $O$, as long as no confusion is present. Consequently, the generator of our C.M.C is given by the following expression:

$$A = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$$

And the probability transition matrix $P(t)$ as:

$$P(t) = e^{At} = \frac{1}{\mu + \lambda} \begin{pmatrix} \mu & \lambda \\ \mu + \lambda & -\mu \end{pmatrix} e^{(\lambda + \mu)t} \begin{pmatrix} \lambda & -\lambda \\ \mu & \mu \end{pmatrix}$$
Deliverable 4.3 HyRiM How to Enhance Perimeter Security Using New Surveillance Technologies

\[
P(t) = I \begin{pmatrix} I & O \\ O & O \end{pmatrix} \begin{pmatrix} \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t} \\ \frac{\mu}{\lambda + \mu} - \frac{\mu}{\lambda + \mu} e^{-(\lambda + \mu)t} \end{pmatrix} O \begin{pmatrix} I \\ O \end{pmatrix}
\]

Let \( \frac{\mu}{\lambda + \mu} = \delta \);

\[
P(t) = I \begin{pmatrix} \delta + (1 - \delta)e^{-(\lambda + \mu)t} & O \\ \delta - \delta e^{-(\lambda + \mu)t} & (1 - \delta) + \delta e^{-(\lambda + \mu)t} \end{pmatrix} O
\]

Besides, \( \pi(t) = (P_I(t), P_O(t)) = \pi(0) \times P(t) \), where \( P_I(t) \) is the probability that subject \( i \) is in state \( I \) (inside area \( j \)) at instant \( t \), and \( P_O(t) \) is the probability that subject \( i \) is in state \( O \) (outside area \( j \)) at \( t \).

As described in Figure 44, the transient regime is changing over time, which governed by the following equations:

- \( (P_I(t), P_O(t)) = \left( \delta + (1 - \delta)e^{-(\lambda + \mu)t}, (1 - \delta) - (1 - \delta)e^{-(\lambda + \mu)t} \right) \) if \( \pi(0) = (1,0) \)
  - If subject \( i \) was in area \( j \) when he was checked at \( t = 0 \).
- \( (P_I(t), P_O(t)) = \left( \delta - \delta e^{-(\lambda + \mu)t}, (1 - \delta) + \delta e^{-(\lambda + \mu)t} \right) \) if \( \pi(0) = (0,1) \)
  - If subject \( i \) was not in area \( j \) when he was checked at \( t = 0 \).

\[\text{(11)}\]

![Figure 44 Transient Regime](image)

At the steady state (independent of time), let denote by \( P_I \) the probability of being in state \( I \); and by \( P_O \) the probability of being in state \( O \). We have:

\[
\begin{cases}
P_I = \delta \\
P_O = 1 - \delta
\end{cases}
\]

Under (H1),

\[P_I = \frac{1}{n} \quad \text{and} \quad P_O = 1 - P_I = \frac{n-1}{n}\]

Which leads to:

\[\lambda = (n - 1)\mu, \quad \text{with} \quad n > 1 \quad \text{according to H3}\]

\[\text{(13)}\]

Let’s focus now on determining when the steady state is reached (\( \Delta t ? \)).

Let \( \varepsilon \) be a small value. We are then looking for the value of \( \Delta t \) such that:

\[P_I(\Delta t) - P_I \leq \varepsilon\]
Which gives us: \( \Delta t \geq \frac{n-1}{\lambda n} \ln \frac{ne}{n-1} \)

To better understand this convergence limit \( \Delta t \), we plotted it as a function of \( \lambda \), for \( n = 2 \) areas, and \( \varepsilon = 0.01 \).

As we can see in Figure 45, the bigger \( \lambda \) is, the faster is the convergence: the bigger \( \lambda \) is, the more rapidly the information “subject \( i \) is in area \( j \)” loses its meaningfulness. This makes perfect sense, since when \( \lambda \) is big, the expected time of stay (expected sojourn time = \( \frac{1}{\lambda} \)) inside area \( j \) is small. The subject \( i \) moves quickly from the area where he was checked, and thus revealing his new position turns back into a guess (\( \frac{1}{n} \)).

**Example:**

Suppose that we have \( n = 5 \). Figure 46 plots the probability of being in area \( j \) as a function of time for different values of \( \lambda \).

![Figure 45 Convergence time (\( \Delta t \)) as a function of \( \lambda \)](image)

![Figure 46 \( P_i(t) \) as a function of time for different values of \( \lambda \)](image)
Of course, the convergence time, $P(t)$, and $P_0(t)$ depend tightly on the knowledge of $\mu$ and $\lambda$. Such information can be easily acquired by an internal attacker (which is our case here), or will require a very long observation from an external observation to be able to infer and estimate these two parameters.

Now, for each subject $i \in S$, we associate $n$ different C.M.C. (one per area $j \in \mathcal{A}$). We will end up with $n \times m$ different C.M.C. As time $t$ ticks, and at any given instant, say $t_0$, corresponding to the occurrence of event $a_{i,j,t_0}$: “subject $i$ is checked in area $j$ at $t_0$”, the $n$ different C.M.C associated to this subject $i$ would be reinitialized as follows:

\[
\begin{align*}
\text{at } t = t_0, & \quad t_i = t - t_0 \\
\text{where } & \quad \left( P_{i,j}(t), P_{0,i,j}(t) \right) = \left( P_{i,j}(t_i = 0), P_{0,i,j}(t_i = 0) \right) = (1, 0) \\
\text{at } t = t_0, & \quad t_i = t - t_0 = 0 \\
\text{where } & \quad \left( P_{i,j'}(t), P_{0,i,j'}(t) \right) = \left( P_{i,j'}(t_i = 0), P_{0,i,j'}(t_i = 0) \right) = (0, 1), \quad \forall j' \neq j 
\end{align*}
\]

As time progresses, we simply apply equation (11) to derive the new probabilities of being in (I), or out (O):

\[
\begin{align*}
\left( P_{i,j}(t), P_{0,i,j}(t) \right) &= \left( \delta_{i,j} + (1 - \delta_{i,j})e^{-(\lambda_{i,j} + \mu_{i,j})t_i}, (1 - \delta_{i,j}) - (1 - \delta_{i,j})e^{-(\lambda_{i,j} + \mu_{i,j})t_i} \right) \\
\left( P_{i,j'}(t), P_{0,i,j'}(t) \right) &= \left( \delta_{i,j} - \delta_{i,j}e^{-(\lambda_{i,j} + \mu_{i,j})t_i}, (1 - \delta_{i,j}) + \delta_{i,j}e^{-(\lambda_{i,j} + \mu_{i,j})t_i} \right), \quad \forall j' \neq j
\end{align*}
\]

After an enough long period of time ($\Delta t_{i,j}$), the steady state is reached, and thus we can simply apply equation (12) to derive the values of $P_{i,j}$ and $P_{0,i,j}$ (independent of time):

\[
\begin{align*}
\left\{ \begin{array}{c}
P_{i,j}(t_i > \Delta t_{i,j}) = P_{i,j} = \delta_{i,j} \\
P_{0,i,j}(t_i > \Delta t_{i,j}) = P_{0,i,j} = 1 - \delta_{i,j}
\end{array} \right.
\end{align*}
\]

And as a final result, we can simply compute our privacy breach based on the probability matrix given by:

\[
P(O) = \begin{pmatrix}
1 & \cdots & \cdots & j & \cdots & \cdots & n \\
\vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\
1 & \cdots & \cdots & P_{i,11}(t_i) & \cdots & \cdots & P_{i,1n}(t_i) \\
\vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\
1 & \cdots & \cdots & P_{l,11}(t_i) & \cdots & \cdots & P_{l,1n}(t_i) \\
\vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\
1 & \cdots & \cdots & P_{m,11}(t_i) & \cdots & \cdots & P_{m,1n}(t_i) \\
\vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\
1 & \cdots & \cdots & P_{l,m1}(t_i) & \cdots & \cdots & P_{l,mn}(t_i) \\
\end{pmatrix}
\]

where $t_i$ is the time variable relative to the origin when subject $i$ was last checked.

As a last part of this section, we must add that this model should be included to the a simulation tool (cf. Section 6.4) to be able to measure this privacy breach over time. Unfortunately, due to the lack of time, this last task would be postponed to the use case scenario in WP5.
7 CONCLUSIONS

The existing surveillance system has problems of inflexibility, misidentification and information leakage. Due to that, unauthorized intrusion could easily happen after long-term attacks in infrastructures. This deliverable analyzes not only physical perimeters but also cyber perimeters and extends the definition of perimeter, including unattended infrastructure, technology populism, outsourcing and human factor. In order to enhance perimeter, some new surveillances are introduced. The new surveillance devices open more flexible and convenient approaches to detect intrusion. In this deliverable, two new surveillance devices, mobile identification check and 3D surveillance camera, are introduced. The integration of on-demand surveillance devices can be done through periodically scheduled operation according to various strategies to detect intrusion and malicious behaviors. The first technology is the 3D surveillance camera. The 3D surveillances camera can provide 3D visual data against noise and depth resolution decay. Additionally, a learning and evaluation framework is proposed to highlight the objects. According to provided data from 3D surveillance camera, the framework tests the reconstruction quality of the data and minimizes the output errors. In this deliverable, in order to give recommendations on the optimal configuration of the video surveillance system, pattern retrieval, pattern clustering, and pattern prediction are applied to obtain risk relevant patterns from the collected video surveilled data. The second studied technology is Mobile ID check. Mobile ID check can effectively verify subjects’ badges and identify intruders. Equipped with security chips and mechanisms, the application and the operating system in the device cannot be changed. Moreover, cryptography technologies are employed against eavesdropping during the communication. Except the CPU itself inside the chip, any other one can neither knows the access to the data nor remove the data. It is easy to realize physical access control with the help of such technology. In order to demonstrate the functionalities of the Mobile ID-check, an experimental simulation model is built. The model adopts three risk metrics including the privacy, the caused damage and the detection rate. The simulation result shows that the provided risk metrics are important elements to find optimal defense strategies. This has been achieved using a multi-criteria and game-theoretical optimization framework developed in WP1. Besides the optimal defense strategy, this framework give indications of worst-case attack strategies per each respective goal as well as the respective optimal loss distributions attained under any behavior of the attacker. However, the surveillance systems also bring privacy breach risks. A privacy model is constructed to evaluate uncertainty in the system based on entropy theory and Markov chain. The privacy model can measure the effects caused by subjects’ original locations and privacy breach caused by frequent ID check missions. The example exhibits that the model can provide a good metric for privacy evaluation.
8 References


ANNEX I: QUESTIONNAIRE FOR ELICITING EMPLOYEES’ PERCEPTION OF PRIVACY

Fast and secure identification of individuals is one of the keys to enforce security. Using a mobile ID check system, a security guard can conduct random identity checks at different locations and at different rates (hourly, every three hours, every day, etc.). Suppose that your employer is willing to set up such a security measure.

If you are to be checked, the guard scans your identification badge (via the camera of a special handheld device) and sends your personal data to a server. In a next step, your personal information (i.e., facial image or fingerprint) is sent back to the device. The guard compares it to your biometrics information and acts accordingly.

During these activities, some of your private data could be exposed (e.g. biometrics, location, etc.). The aim of this survey is to collect some empirical evidences on the degree to which you, as an employee, feel uncomfortable or being watched. The potential privacy infringements measure will be a qualitative score (ranging from 0 = “totally privacy friendly” up to 10=” infringing your privacy completely”).

1. Does your working area have an access restriction, so that only authorized employees can enter?
   □ Yes          □ No          □ Unknown

2. What do you feel is the acceptable number of badge checks during a day?
   □ ≤ 2          □ 3          □ 4          □ 5
   □ 6          □ 7          □ 8          □ ≥ 9

3. How would you rate the different ID check frequencies with regard to the imposed privacy infringement? In the following table please mark with an X for each number of checks per day, from line 1 check/day to greater than 9 checks/day, what you do feel is acceptable on the sliding scale from totally privacy friendly, 0, to infringing your privacy completely, 10.

<table>
<thead>
<tr>
<th>Number of Checks per day</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
</tr>
<tr>
<td>≥ 9</td>
<td></td>
</tr>
</tbody>
</table>
## ANNEX II: PAYOFF STRUCTURES FOR GOALS DAMAGE, DETECT RATE, AND PRIVACY

<table>
<thead>
<tr>
<th></th>
<th>D-NG5F7H8LF</th>
<th>D-NG3F7H8LF</th>
<th>D-NG1F7H8LF</th>
<th>D-NG5F7TR</th>
<th>D-NG3F7TR</th>
<th>D-NG1F7TR</th>
</tr>
</thead>
<tbody>
<tr>
<td>damage</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>loss(1)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>loss(2)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>loss(3)</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>loss(4)</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>loss(5)</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Figure 47 Loss distribution (Payoff structure) for goal “Damage”
Figure 48 Loss distribution (Payoff structure) for goal “Detection Rate”
Figure 49 Loss distribution (Payoff structure) for goal “Privacy”