SINGIDUNUM UNIVERSITY
Department for postgraduate studies

Doctoral Dissertation

CONTRIBUTION TO INFORMATION SECURITY CONTINUOUS AUDIT IN CLOUD-NATIVE ENVIRONMENTS

Supervisor:

Prof. dr. Milan Milosavljević

Candidate:

Meiran Galis

Belgrade, 2023.

Supervisor:

Prof. dr Milan Milosavljević, Full Professor
Singidunum University, Belgrade

Committee members:

Prof. dr Milan Milosavljević, Full Professor
Singidunum University, Belgrade

Prof. dr Mladen Veinović, Full Professor
Singidunum University, Belgrade

dr Tomislav Unkašević, Associate Research Professor, co-supervisor
VLATACOM Institute, Belgrade

Date of defense: _____

Acknowledgement

I would like to thank my supervisor Prof. Dr. Milan Milosavljević and the co-supervisor Dr. Tomislav Unkašević for their continued support, valuable advice and cooperation during the implementation of this research and the preparation of this doctoral dissertation. I owe a great debt of gratitude to the VLATACOM Institute for providing me with the opportunity, resources and time to conduct research, as well as great support during the preparation of the doctoral dissertation. I would also like to thank all members of the cryptology depatment who, with their knowledge, experience and support, provided me with a stimulating environment for work and research.

Abstract

In the digitalized world and Cyberspace, as symbiotic community of men and machines, Cloud computing technologies and digital services based on them have important role in everyday life and business processes. From an information security standpoint, a whole range of security challenges arise, starting with security goals and security architecture through their operationalization and implementation. This is particularly reflective of the information security audit as part of the audit of information systems. In terms of information security cryptographic algorithms and cryptographic protocols are significantly standardized and support the approach of continuous external audit and improvement of the security of the subject information system. On the other hand, all of these solutions involve the use of cryptographic parameters created appropriately and under certain conditions. This audit segment requires specialist knowledge and the ability to assess the adequacy of the procedures applied. Contrary to cryptographic algorithms and protocols in this segment, there is no generally accepted standardization. This research is an attempt to develop a method that would be reliable in theoretical terms and proofs and also independent of trusted third parties. Such a method would significantly improve the possibilities of continuous revision in this segment and information security in the systematic sense.Suggested method is based on biometrical data, recorded electro-encephalography signals, randomness extraction from stochastic processes with non-maximal entropy and methods for transformation stochastic sequences for their uncertainty improvement. It is shown that it is possible to obtain truly random sequence sheared between participants in the protocol using communication over publically available authenticated communication channel. An unauthorized observer is able to collect all exchanged messages but in information sense cannot collect enough data to reconstruct established content between the two entities, and this can be theoretically proven. In the process, there is no trusted third party that entities must trust and have control over them and their communication, implying autonomy in setting end-to-end protection.

Contents

Chapter 1

Introduction

Modern information technologies and globalization have dramatically changed today's world of computer networks, which are characterized by a high degree of integration of various electronic services. As the number of Internet services and new users of services increases every day, the amount and value of information exchanged increases. Information exchanged over networks and storage on networked memory locations may be compromised if not adequately protected.

The growth of communication and network technologies, coupled with advancements in the design and deployment of microprocessor devices, has enabled the seamless exchange of information among various environments such as sensors, devices, and information systems. This has facilitated the development of intelligent systems capable of overseeing and managing intricate processes, as well as providing a range of services. For that purposes the Internet infrastructure and protocols are utilized and used to create a space of mutually connected entities, like Wireless Sensor Networks (WSN) and Internet of Things (IoT), or remote information storage, their processing and based on them remote services (cloud computing). The advancement in technology plays a pivotal role in advancing various technological and life processes, giving rise to smart cities, autonomous vehicles, robotization, and intelligent robot behavior. In this context, information security holds a crucial role, as any compromise in the integrity and privacy of data within this integrated world could result in severe consequences, possibly leading to a widespread disaster. Hence, besides the security mechanisms embedded in Internet protocols, additional safeguards are integrated into devices and systems to prevent unintended behavior.

Furthermore, a multitude of such devices and services operate in real-time scenarios, requiring the defined security mechanisms to seamlessly align with system behavior without causing disruptions. These mechanisms must be designed for easy implementation in both hardware and software, ensuring their efficiency so that users can comfortably utilize the offered services.

In the digitalized world and Cyberspace, as symbiotic community of men and machines, Cloud computing technologies and digital services based on them have important role in everyday life and business processes.

Increasingly enterprises are incorporating cloud-based applications into their regular business operations. In order to protect their confidential data, if required by legal commitments or if by regulations, they need to verify that their IT vendors follow relevant security standards and privacy regulations.

The rapidly evolving cloud utilization of corporations migrating to the cloud or new technology companies (start-ups) has led enterprises to require security certifications based on independent external auditors.

The speed of change requires companies to shift their focus away from manual testing of risks and controls to critical and strategic risks. Automating repetitive tasks that are considered lower value by organizations, such as compliance testing, is often a priority.

By outsourcing routine audit tasks to third parties, internal resources can be used more strategically and beneficial to the organization. But, outsourcing Information security audit and automatization of it imply proven security methods usage, auditable and secure methods of their implementation.

From an information security standpoint, a whole range of security challenges arise, starting with security goals and security architecture through their operationalization and implementation. This is particularly reflective of the information security audit as part of the audit of information systems. In terms of information security cryptographic algorithms and cryptographic protocols are significantly standardized and support the approach of continuous external audit and improvement of the security of the subject information system. On the other hand, all of these solutions involve the use of cryptographic parameters created appropriately and under certain conditions. This audit segment requires specialist knowledge and the ability to assess the adequacy of the procedures applied. Contrary to cryptographic algorithms and protocols in this segment, there is no generally accepted standardization. This research is an attempt to develop a method that would be reliable in theoretical terms and proofs and also independent of trusted third parties. Such a method would significantly improve the possibilities of continuous

revision in this segment and information security in the systematic sense.

Internal and external audits in this sense can be considered local because they are a means of detecting local characteristics of companies and their business processes (companies in and of themselves or between several companies).

A third-party audit can be considered global in some way because it is usually a check on the fulfilment of conditions in accordance with legal regulations. This is particularly prominent in the area of information security and may, for example, apply to regulations that require a level of security to be reached in accordance with state law, for example in the area of certification bodies for issuing qualified electronic certificates, electronic identification documents and machine-readable travel documents. Revision of such systems as mentioned above is not at all easy due to the necessity of specific knowledge regarding the processes of generating cryptographic parameters. Despite very serious and exhaustive testing, some security weaknesses were not detected in time and caused problems in the functioning of such systems. This fact illustrates the importance of reliably generating cryptographic parameters and applying such methods in the implementation of information systems security mechanisms.

The aim of this research is to define methods for generating cryptographic parameters between the two entities in communication in such a way that:

1. Generated cryptographic parameters in statistical terms have theoretically proven characteristics of a truly random sequence.

2. An unauthorized observer in information sense cannot collect enough data to reconstruct established content between the two entities, and this can be theoretically proven.

3. In the process, there is no trusted third party that entities must trust and have control over them and their communication.

The method with these characteristics has wide application in the field of information security and continuous audit in the field of cloud technologies and services for the following reasons:

1. Proving the randomness of obtained cryptographic parameters meets fundamental cryptographic requirements and, given the theoretical support, does not require specialist knowledge during the audit.

2. The provable inability to reconstruct the resulting parameters consequently implies the security of the process and recommends it for use in defining security solutions in the field of information security. Theoretical proved characteristics eliminates the need for specialist knowledge during continuous audit.

3. Bearing in mind the previous two points, it concloudes that the review of information security can, thanks to this method, be reduced to an audit of its implementation and standardized security methods. This makes things much easier with the outsourced audit of security solutions because the technologies for safe implementation of software solutions are well developed and their audit is straightforward.

The method developed during this research is based on the extraction of electro encephalographic signals from different people induced by the same visual content. The method proposed by this research meets the abovementioned requirements.

Structure of the dissertation is as follows.

First Chapter will present the motivation for this research, problem and subject of the research and overview of the dissertation contents.

Second Chapter presents in brief cloud computing technologies, their taxonomy and key features. Also, relationship of cloud based information systems regarding continuous auditing and information security is described.

In that context brief introduction into information system auditing will be presented.

Third Chapter of the dissertation describes basic information security principles and cryptographic basis of information security with accent on the randomness of cryptographic parameters. Connection of the cryptographic parameters generation and continuous auditing is described.

Forth Chapter, contains description of Information theory approach to cryptographic parameters generation and its formal model

Fifth Chapter is the central part of the dissertation. In this part Information theory method for common random string establishment based on collected EEG signals aroused by the same mental stimulus between communication participants. Exact method will be presented in this part with theoretical arguments for its correctness. Correctness is based on information-theoretic and statistical analysis. Set of parameters which influence length of obtained common random strings is identified and strategies on choosing

parameters values are defined. Results of extensive experimental analysis are presented in this chapter.

Sixth Chapter is closing part of the dissertation. In this part summation of the research is presented with achieved goals and results. According to the achieved results contribution of the research is described with potential applications and directions of further research is listed.

The proposed approach introduces a novel approach for common random string establishment between communicating participants. Method offers number of benefits for application in information security solutions. Beside the direct benefits for Information security additional benefits lay in the field of continuous auditing of security in cloud computing environment. The benefit is reflected in the fact that one such method, formally based on the arguable characteristics, enables an audit based on formal and automated procedures

Chapter 2

Cloud Computing Technology Overview

The ideas of distributed computation and collaborative use of computing resources have their roots since the mid-1970s and experience their full affirmation with the development of information, communication and computing technologies through the cloud computing paradigm, [1], [2]. This approach completely reshapes business processes from the point of view of computer and communication resources, their organization, access and management. This technology, often referred to as the "cloud", offers flexible and scalable access to information communications, computing and functional infrastructure through access to various resources on demand through the use of network communication infrastructure. The technology itself is characterized by access models, types of services and usage models.

2.1 Service models

2.1.1 Infrastructure as a Service

Infrastructure as a Service (IaaS) is one of the key cloud models that provide basic infrastructure and computing resources over the Internet. IaaS allows users to virtualize and use hardware in the form of virtual machines, mass memory space and storage, without the need to manage physical hardware. Basic features of IaaS are:

1. Basic infrastructure: IaaS provides basic infrastructure through virtual

machines (VMs), physical and virtual servers, memory for long-term and short term data storage (disks, RAM memories) and network infrastructure (virtual networks, firewalls, management of levels of resource utilization and their load).

2. Independent control: Users have a high level of control over virtual machines and resources. They can create, configure, and manage VMs according to their needs. This allows for adaptability and scalability.

3. Pay by consumption: IaaS often uses a "post-paid" model, meaning that users only pay for resources to the extent that they have used them. This allows cost optimization and scalability to meet needs.

4. Responsibility for operating system and applications: Users are responsible operational use and maintenance of the operating system and applications that run on virtual machines. This encompasses operating system and applications maintenance and their security patches.

5. Elasticity and scalability: IaaS enable quick and easy resource scalability. Users can add or remove virtual machines according to needs, which is especially useful for applications with variable load and ununified resource needs.

6. High availability and security: IaaS providers typically provide a high degree of availability and various security options such as data replication, redundant storage, and advanced system security mechanisms.

7. Data layout according to geographical accessibility: IaaS enables providers to deploy their resources around the world, in geographically dispersed data storage centers and close to the user, which improves accessibility and reduces response times to customer requests in different regions.

In this context widely known providers are Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP), IBM Cloud, and many others. IaaS has significance for organizations that want more control over their IT infrastructure but do not want to deal with the physical management of hardware and data centers.

2.1.2 Platform as a Service

Platform as a Service (PaaS) is a service paradigm which assumes a platform and programming environment, execution and management of applications without taking a care about the basic computer infrastructure and hardware. PaaS services incloude tools and services for application development, integration, deployment, and scale-up. This type of service simplifies application development and allows developers to concentrate to coding and functionality in place of infrastructure maintenance and management. Key features of PaaS:

1. Application Development Tools: PaaS platforms offer a range of development tools, inclouding environments for working in different programming languages, context-sensitive editors, and libraries, to facilitate application development. Developers can write, test and correct detected bugs in their code within the platform.

2. Software packages and integration: PaaS often allows the use of commercial software packages that help integrate data, messaging, and communication between different components of the application. This simplifies the process of shaping the functionality of the software and connecting different functionalities.

3. Scalability: PaaS solutions provide the ability to adaptively adjust the scope of resource usage, up or down, according to the load on the application itself, if designed that way. This allows you to manage the costs of using the created software.

4. Resource management: PaaS abstracts the core infrastructure, so developers do not have to manage resources but hire them according to their needs and the scope that the service provides.

5. Database and storage services: PaaS typically incloudes databases and storage solutions that facilitate data management within applications. These services often incloude backups, data replication, and data management functions.

6. Security and compliance: PaaS often incloude security methods such as authentication, encryption, and access control to secure applications and data. Many of PaaS' offerings are also in line with industry standards and regulations.

7. Monitoring and management: PaaS platforms typically provide tools for monitoring and managing applications to monitor the performance and correctness of applications. This incloudes event logs during application operation and performance analysis to help developers and operators identify possible problems and congestion for optimization application performance.

8. Cost efficiency: PaaS can be financially efficient because it reduces the need for organizations to buy, administer and manage their devices and infrastructure. Users are charged based on their actual use, which makes it easier to control costs.

9. Developer productivity: PaaS platforms are designed to increase developer productivity by simplifying development processes and reducing time spent on infrastructure-related tasks. This allows for faster development and application cycles.

10. Multi-Tenancy: PaaS platforms are often designed to support multiple users or tenants, providing isolation between different applications or users on the same infrastructure.

11. Vendor Lock: One possible disadvantage of PaaS is the risk of the impossibility of migrating applications to another platform because they are developed using platform-specific tools and are vendor locked in.

Examples of well-known PaaS providers are Microsoft Azure App Service, Google App Engine, Heroku, AWS Elastic Beanstalk, and Red Hat OpenShift.

PaaS is a valuable option for developers and organizations that want to focus on application development while unloading infrastructure management tasks to a cloud service provider. It is suitable for web and mobile application development, micro service architecture, and other cloud approaches.

2.1.3 Software as a Service

Software as a Service (SaaS) allows software applications customer usage by the Internet as communication technology. Users can execute these applications using browser client interface, removing the need for locally installed

ones. Popular examples of SaaS incloude email services such as Gmail and office productivity packages such as Microsoft 365. Software as a Service (SaaS) is a cloud computing service paradigm that provides access to software by user registration and payment for the application usage in regular time slots, opposite to buying software and installing it on its own hardware. Here are the key characteristics and details of SaaS:

1. Accessibility: SaaS applications are accessible via the internet, allowing users to access them from any place with available Internet access. This accessibility makes it convenient for remote work and collaboration.

2. Subscription-Based: SaaS is typically offered on a subscription basis, where users pay a recurring fee in regular time slots to access the software. This subscription model often incloudes updates, support, and maintenance.

3. No Installation or Maintenance: Software is installed on provider infrastructure and users do not need to install or maintain it on their local devices or servers. All maintenance, inclouding updates, patches, and security, is managed by the SaaS provider.

4. Multi-Tenancy: SaaS applications are typically designed to support multiple customers (tenants) on the same infrastructure, with data and access isolation to ensure data privacy and security.

5. Automatic Updates: SaaS providers regularly update their software to improve features, security, and performance. Users automatically receive these updates without having to manage the process themselves.

6. Scalability: SaaS applications are often designed easy scalability, so they can fit in with a growing number of users or changing workloads. Users can often coordinate their subscription level to match their needs.

7. Data Accessibility: SaaS applications store user data in the cloud, allowing access to them from any place with internet connection. This facilitates data sharing and collaboration among users.

8. Collaboration and Sharing: Many SaaS applications are designed for collaboration, allowing multiple users to work on the same document or project in real time. This is particularly beneficial for remote teams.

9. Integration: SaaS providers often offer APIs (Application Programming Interfaces) to allow integration with other software and services, enabling users to connect SaaS applications with their existing systems.

10. Reduced IT Overhead: SaaS eliminates the effort made by an organization to manage the underlying infrastructure, reducing IT overhead costs related to hardware, software installation, and ongoing maintenance.

11. Data Security: SaaS providers typically invest in strong, cutting edge, security measures to protect user data. This incloudes cryptographic protection, access control methods, and compliance with data protection regulations.

12. Pay-as-You-Go Pricing: Some SaaS providers offer a post-paid pricing model, where organizations are billed based on actual usage. This can be cost-effective for businesses with varying workloads.

13. Vendor Lock-In: Organizations should be aware of the potential for vendor lock-in when using SaaS solutions, as migrating data and processes to another provider or on-premises can be complex and costly.

Examples of popular SaaS applications incloude Microsoft 365 (formerly Office 365), Google Workspace (formerly G Suite), Salesforce, Dropbox, Zoom, and Slack. SaaS is suitable for a wide range of business applications, inclouding office productivity, customer relationship management (CRM), project management, collaboration, communication, and more. It is particularly beneficial for businesses looking to reduce IT infrastructure and maintenance costs while gaining access to up-to-date software and features.

2.2 Application models

The models of organization of resources, methods of access and management by cloud computing service providers at a logical level differ mainly in visibility and methods of access to services and availability of resources.

2.2.1 Public cloud

In this model, resources are located on infrastructure that is accessible to more organizations. It is an economically viable solution, but it induces security issues and compliance with information security legislation. A public cloud is where infrastructure and services are available to the public, usually on the basis of a subscription in the form of a certain financial amount addressed to the cloud service provider. Service providers are implemented in data centers that can be accessed via the Internet. The main characteristics of the public cloud are:

1. Accessibility: Public cloud services are available to anyone with an internet connection, making them accessible to a wide range of users and organizations, inclouding individuals, startups, and businesses.

2. Shared resources: Public clouds are multi-tenant environments, where resources are shared among multiple customers or tenants. This sharing enables cost efficiency and resource optimization.

3. Pay-As-You-Go pricing: Public cloud providers typically offer a pricing model based on payment in further email or subscription. Users pay for the resources and services they consume, enabling cost flexibility and scalability.

4. Scalability: Public cloud services are designed to easily adapt to the required scope of use. As needed, users can dynamically allocate additional resources to accommodate increased workload or traffic growth. This scalability promotes agility and cost efficiency.

5. A variety of services: Service offered by the cloud service providers incloudes large pallet of services, inclouding Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). Services usually cover computing, storage, databases, networking, machine learning, analytic and more.

6. Global data centers: Cloud service providers manage resource centers in different, possibly distanced regions and countries, enabling users to deploy resources and services in locations that are geographically close to the user or comply with specific data storage and storage requirements.

7. Security and compliance: Public cloud provider's are making great efforts to adequate security measures to protect users' data. They typically offer a large scale of security features, such as cryptographic protection, access controls, identity management, as well as compliance certifications (e.g., GDPR, HIPAA).

8. Reliability and availability: Public clouds are designed for high availability and reliability. Frequently, suppliers are prepared to provide service level contracts (SLAs) that guarantee a certain degree of quality for the services they provide. They use redundancy, failure mechanisms, and data replication to reduce downtime.

9. Elasticity: Public clouds allow rapid provision of resources as a reaction to the increased customer needs. This elasticity makes it easy to adapt to changing business needs without significant investments in advance.

10. Managed Services: Public cloud providers offer managed services that solve a variety of tasks, such as automated backups, patch management, and monitoring. This reduces the operational load on users and allow them focusing on application development.

11. Developer Tools: Public cloud vendors have at their disposal variety of developer tools and APIs to create, implement and maintenance applications and services. These tools facilitate integration and automation.

12. Global Reach: Public clouds have a global presence, enabling organizations to expand and serve customers in different regions without having to set up their own infrastructure at each location.

13. Community and Marketplace: Some public cloud providers offer community forums and marketplaces where users can share third-party knowledge, tools, and applications.

Popular public cloud providers incloude Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP), IBM Cloud, and Oracle Cloud. Public cloud services are used for a variety of purposes, inclouding web hosting, application development, data storage and processing, big data analytic, machine learning, and more. They are a key enabler of digital transformation and cloud-based computing for businesses of all sizes.

2.2.2 Private cloud

Private cloud is a dedicated cloud infrastructure for an organization that provides greater control and security. It is appropriate for organizations with strict information security and privacy requirements. Private cloud is a model of deploying cloud technology where infrastructure and services are provided and maintained for the exclusive use of an organization. Usually they are hosted in the dedicated data center. Private clouds offer greater control, security and adaptation compared to public clouds.

The key features and details of private clouds are:

1. Exclusive use: Private clouds are used by a single organization, meaning they have exclusive access to infrastructure and services. This offers more control over resources and data.

2. Security and privacy: Private clouds are often perceived as safer than public clouds. Organizations have control over security measures, access policies and data protection. This is especially important for industries with strict compliance requirements.

3. Customization: Private clouds provide the flexibility to customize the cloud environment to specific business needs. Organizations can configure and optimize infrastructure and services according to their unique requirements.

4. Data control: With the private cloud, organizations have full control over their data. Data remains in the organization's data center or hosted in a dedicated environment, reducing concerns about data exposure or unauthorized access.

5. Regulatory compliance: Private clouds are suitable for industries with regulatory and compliance conditions, such as healthcare (HIPAA), finance (PCI DSS), and government (FISMA). Organizations can apply the necessary controls to meet these standards.

6. Predictable performance: Private clouds offer predictable performance because resources are not shared with other tenants. This makes private clouds a good choice for applications with strict performance requirements.

7. Insulation: Private clouds provide strong insulation, ensuring that the activities of one tenant do not affect the performance or safety of other tenants. This isolation can be crucial for the implementation of a critical mission.

8. Hybrid cloud integration: Private clouds can be integrated with public clouds to create a hybrid cloud environment. This approach assumes to take best features from both models, scalability of public clouds and high information security for sensitive data and critical workloads in the private cloud.

9. High availability: Private clouds are designed for high availability, often with redundant infrastructure and failure mechanisms to reduce downtime.

10. Comprehensive management: Companies or other users have full control over the management of their private cloud, inclouding resource provision, monitoring and maintenance. This can be an advantage for IT teams that want to maintain control.

11. Cost predictability: Private clouds often have predictable cost structures, as organizations can allocate resources as needed without the variable costs associated with public cloud payment models.

12. Infrastructure investment: Setting up and maintaining a private cloud usually requires a significant up-front investment in hardware, software, and IT resources. However, this can be cost-effective for organizations with long-term cloud needs.

13. Simplified networking: Private clouds often offer simpler network configurations, as organizations have control over the networking infrastructure, making it easier to implement certain network policies and connectivity requirements.

Widely used private cloud solutions encompasses, among the other companies, OpenStack, VMware vCloud, and Microsoft Azure Stack. Private clouds are typically used by large businesses, government organizations, and industries with stringent data security and compliance requirements. They provide a controlled and safe environment to run critical workloads while offering flexibility and customization.

2.2.3 Hybrid cloud

Hybrid cloud model assumes usage of best characteristic of public and private cloud model utilizing data and application transfer between them and that movement is stealth for the customer. This model provides flexibility and is used by organizations that need to balance cost and safety. This approach offers a range of benefits, inclouding flexibility, scalability and data sovereignty. Here are the key features and details of the hybrid cloud:

1. A fusion of public and private clouds: Hybrid clouds consist of both public and private cloud components, with data and applications seamlessly integrated and shared between the two environments.

2. Flexibility: Organizations can leverage the public cloud for scalability and cost efficiency, while maintaining control over confidential information data and critical loads in the private cloud. This flexibility allows for a more efficient allocation of resources.

3. Scalability: Hybrid clouds allow organizations dynamic allocation of the resources up or down as necessary, using public cloud resources during peak demand and private cloud resources for constant load.

4. Data and application portability: Hybrid cloud architectures allow data and applications to move between used cloud technology models. This portability can be valuable for cost optimization, redundancy, and disaster recovery.

5. Data sovereignty and compliance: Private clouds are often used for data that must meet regulatory and compliance specific requirements, while public clouds provide flexibility for data storage and processing globally. Hybrid clouds help organizations meet both local data residency and global processing needs.

6. Security and control: Private clouds offer an increased level of control and security of sensitive data and applications. Organizations can maintain their own security policies and controls in the private cloud while benefiting from public cloud services.

7. Cost optimization: Organizations can optimize costs by using public cloud resources when it makes sense, like periods of high demand, and

then return workloads to the private cloud when resources are underutilized.

8. High availability and disaster recovery: Hybrid cloud technology permit high level availability and disaster recovery strategies by redundant storage for sensitive data and applications between public and private clouds. This ensures that data and applications remain available in the event of failure.

9. Consistent management: Hybrid cloud management platforms provide a unified interface for managing both public and private cloud resources, simplifying administration, and reducing operational complexity.

10. Integration and interoperability: Hybrid cloud architectures often rely on APIs and connectors to facilitate seamless integration between public and private cloud components, enabling data and application synchronization.

11. Redundancy: By deploying workloads and data across public and private clouds, organizations can achieve redundancy and reduce the risk of falling in the event of failure or interruption in a single cloud environment.

12. Resource optimization: Organizations can allocate resources to fit the necessary levels of their workloads, using public cloud resources when additional capacity and private cloud resources are needed for mission-critical loads.

13. Geographic allocation: Hybrid clouds allow organizations to deploy resources and applications across a variety of regions and data centers, bringing services closer to end users to improve performance and reduce latency.

14. Popular hybrid cloud solutions and vendors incloude AWS Outposts, Microsoft Azure Arc, Google Anthos and VMware Cloud. Hybrid cloud environments are typically used by businesses looking to balance the benefits of public cloud services with the control and security of private clouds. They are especially suitable for organizations with complex IT requirements and a variety of workloads.

2.2.4 Multi-Cloud model

Use the services of multiple cloud vendors. It helps avoid vendor lockdown, improves redundancy, and provides options for using specialized services from different vendors. Multi-cloud is a cloud computing strategy which entails making use of a number of different cloud providers in order to host and administer heterogeneous components of an organization's information technology infrastructure. This approach offers flexibility, redundancy and the ability to choose the best services from different service providers. Here are the key features and details of the multi-cloud strategy:

1. Multiple cloud providers: Multiple clouds assume usage of several, more than one, service provider's services to achieve different needs or to host different components of an organization's infrastructure. These providers can be a mix of public, private or hybrid clouds.

2. Diversity of services: Organizations choose cloud providers based on the unique features and services they offer. This allows them to choose the best supplier for each particular item of use or workload.

3. Flexibility and avoiding vendor locks: Multi-cloud strategies provide flexibility and reduce the risk of vendor lockdown. If one cloud provider does not meet an organization's requirements, it can use another provider without costly migration.

4. Redundancy and high availability: By distributing loads across multiple cloud providers, organizations can achieve redundancy and high availability. This reduces the risk of falling working hours in the event of cancellation or interruption of work with a single service provider.

5. Cost optimization: Multi-cloud enables organizations to optimize costs by choosing the most valued cloud providers and services for their workloads. This can lead to significant cost savings.

6. Multi-cloud enables organizations to deploy resources and applications in different regions and data centers, bringing services closer to end users to improve performance and reduce latency.

7. Data residency and compliance: Organizations can choose cloud vendors that comply with specific data residency requirements and regula-

tory compliance standards, ensuring that data is stored in appropriate geographic regions.

8. Multi-cloud allows organizations to allocate resources to suit the specific needs of their workloads. For example, public cloud resources can be used for scalability, while private cloud resources can host sensitive data.

9. Integration and interoperability: Multi-cloud environments require integration and interoperability between different cloud providers, often through the use of APIs and connectors. This facilitates seamless data and synchronization of applications.

10. Data and application portability: Multi-cloud environments enable the movement of data and applications between cloud vendors, providing flexibility in terms of resource allocation and redundancy.

11. Complexity and management challenges: Managing multiple cloud suppliers can introduce complexity and require careful orchestration. Specialized management tools or platforms may be necessary to simplify administration.

12. Security and identity management: Multi-cloud security strategies need to consider identity and access management across different cloud providers, providing consistent security policies and controls.

13. Monitoring and Performance Optimization: Performance monitoring and optimization tools are essential in a multi-cloud environment to ensure that workloads work efficiently and efficiently across a variety of vendors.

14. Skills and expertise: Organizations may need a variety of skills and expertise to effectively manage and secure resources across multiple cloud vendors.

15. Hybrid and Multi-Cloud Distinctions: While hybrid clouds combine public and private clouds, multi-cloud extends beyond these deployment models to involve multiple providers, which can involve combinations of public, private and hybrid clouds.

Popular tools and platforms used in multi-cloud environments incloude Kubernetes for container orchestration, cloud management platforms such as CloudHealth and RightScale, and infrastructure-as-code (IaC) automation and resource provision solutions.

Multi-cloud strategies are typically used by organizations that want to maximize flexibility, reduce risk, optimize costs, and maintain control over their cloud resources. They are suitable for organizations with different workloads and specific requirements for compliance or data stay.

2.3 Key features of Cloud computing technology:

Cloud computing technology is a transformative force that has redefined the way computing resources are accessed and managed. Its flexibility, scalability and cost efficiency make it an invaluable asset for organizations and individuals who want to harness the power of the cloud for a multitude of applications and services. Understanding the components and benefits of cloud computing is essential for making informed decisions and maximizing the benefits it offers. Main features and benefits are:

- Scalability: Cloud resources can be easily adjusted up or or down to meet demand. This agility is especially valuable for businesses with a fluctuating workload.
- Cost-Efficiency: Cloud computing eliminates the need for extensive on premise infrastructure, reducing capital costs. Users only pay for the resources they spend.
- Flexibility: Cloud offers a wide range of services and deployment models, allowing users to choose the best that suits their needs.
- Accessibility: Resources can be accessed from any place which make possible to access Internet, enabling remote work, collaboration and global accessibility.
- Resource Pooling: Cloud providers use models with multiple tenants, where resources are shared among multiple users. This pooling of resources leads to cost savings and efficient utilization.

- Self-service: Users can provide and manage resources without the need for extensive technical expertise, thanks to interfaces and tools for managing custom users.

Cloud computing is employed in various use cases, such as:

- Data storage and backup: Cloud storage services such as Amazon S3 and Google Cloud Storage offer secure and scalable data storage solutions.
- Web Hosting: Cloud-based web hosting services provide high availability and scalability for websites and web applications.
- Big Data Analytics: Cloud platforms such as AWS and Azure offer tools and services for processing and analyzing large datasets.
- Machine learning and AI: Cloud platforms provide computing power and tools necessary for machine learning and artificial intelligence applications.
- Disaster recovery: Cloud-based disaster recovery solutions ensure backup and data redundancy in the event of system failures or disasters.

2.4 Continuous auditing of Cloud computing systems

Continuous auditing in the context of cloud computing systems is a proactive and iterative approach to monitoring, assessing, and making certain that the infrastructure, apps, and data stored in the cloud are secure and in conformity with regulations. This practice is essential in the dynamic and ever-evolving landscape of cloud computing, where traditional audit methodologies may fall short in providing real-time insights into security and compliance postures. Next we will consider relationship between continuous auditing and cloud computing, key principles, challenges, and examples.

Key Principles of Continuous Auditing in Cloud Computing:

– Real-time Monitoring

Real-time Monitoring assumes continuous auditing which involves real-time monitoring of cloud resources and activities to promptly detect

and respond to security incidents or deviations from compliance standards. For that cloud-native monitoring tools are utilized to track user access, configuration changes, and network activities in real-time. Alerts can be triggered for suspicious behavior; responsible entities are alarmed enabling immediate investigation and response.

- Automation and Orchestration
  Automation is integral to continuous auditing, enabling the regular and systematic examination of cloud configurations, security controls, and compliance policies. This is important part because the reaction latency in the potentially incident situations may be inappropriately long in the case of only human system monitoring comparison. Automated scripts and tools can regularly scan cloud environments for misconfigurations, adherence to security best practices, and compliance with industry standards. This ensures consistency and efficiency in the auditing process.

- Scalability and Flexibility
  Continuous auditing should be scalable and adaptable to the dynamic nature of cloud environments, where resources can be rapidly provisioned, modified, or decommissioned. For example, implementing audit scripts and policies that can scale with the organization's cloud usage, accommodating changes in infrastructure assuming hardware, software, and data storage. This ensures that the auditing process remains effective as the cloud environment evolves.

- Integration with DevOps Practices
  Principle: Continuous auditing aligns with DevOps principles, integrating security checks seamlessly into the development and deployment pipelines. It is possible to achieve this by security checks implementation as code within the CI/CD (Continuous Integration/Continuous Deployment) pipeline to recognize, mark and act regarding detected security vulnerabilities and misconfigurations early in the development process.

- Comprehensive Data Collection
  Continuous auditing requires the collection of comprehensive data on cloud activities, configurations, and events to facilitate thorough analysis. For example, logging and monitoring data related to user access,

system changes, API calls, and network traffic. Aggregating and analyzing this data provides insights into the overall security posture and potential risks.

Continuous auditing of Cloud computing systems brings with it some non-trivial challenges such as:

- Diversity of Cloud Services
  Cloud environments often comprises a diverse set of services, each with its own configurations and security considerations. For these reasons it is necessary to tailor continuous auditing processes to the specific characteristics of each cloud service with auditing procedures giving more appropriate and reliable results. This may involve using service-specific auditing tools and policies.

- Ephemeral Nature of Resources
  Cloud resources can be provisioned and decommissioned rapidly, leading to challenges in tracking and auditing ephemeral resources. Usual method for mitigation this type of problems is to leverage automated discovery mechanisms to identify and audit newly provisioned resources. This assumes tools and tagging strategies to categorize resources and track their purpose.

- Security and Compliance Variability
  Different cloud service providers may have varying security controls and compliance standards, requiring adaptable auditing processes. In this type situation it is necessary to customize continuous auditing scripts and policies based on the specific features and security controls offered by each cloud provider. It is necessary to update regularly auditing practices to align with evolving provider capabilities.

- Data Privacy Concerns
  Continuous auditing involves collecting and analyzing sensitive data, raising privacy concerns. Mitigation: Implement anonymization or encryption techniques when handling sensitive data during the auditing process. Ensure compliance with data protection regulations and obtain necessary consents.

2.5 Relationship of cloud computing and information security

The relationship between information security and cloud computing is integral and complex, as the adoption of cloud technologies transforms the traditional IT landscape, [2]. Cloud computing offers scalability, flexibility, and cost-effectiveness, but it also introduces new challenges and considerations for information security. In the next part we will consider the multifaceted relationship between information security and cloud computing.

– Data Security

  Regarding the data security important characteristic is that in a cloud environment, data is stored on remote servers managed by third-party providers. This raises concerns about the security and privacy of sensitive information in the sense of rights to access to them and prevention of unauthorized their content disclosure and usage. Prevention is based on cryptographic methods. Data ciphering both in transmission and during their storage is crucial. Cloud providers typically offer encryption services, and organizations should also manage encryption keys securely. Access controls and identity management help ensure that only authorized users can access sensitive data.

– Identity and Access Management (IAM)

  As we formerly mentioned IAM is very important regarding information security. Cloud computing involves multiple users accessing resources from various locations. Managing and securing identities and access becomes critical. This indicates implementation of robust IAM policies, inclouding strongly secure identity management system, multifactor authentication, to verify user identities. Role-based access control (RBAC) ensures that users have appropriate permissions based on their roles, mitigating the risk of unauthorized access. Here is important to note that security of all mentioned methods depends on the identity management system because the evidence of who, when and what action performed is crucial for the analytics of incident and acting appropriately.

– Compliance and Legal Considerations

  Cloud computing providers usually use data deployed in geographically

different areas. Different regions and industries have specific regulations regarding data storage, processing, and transmission. Ensuring compliance in the cloud can be complex. From those reasons cloud service providers often adhere to industry standards and compliance certifications. However, organizations must understand and ensure compliance with relevant regulations. This incloudes contractual agreements with cloud providers and regular audits to verify compliance.

- Shared Responsibility Model

  Reliability and availability of the cloud computing services depends on cloud computing service provider and behavior and acting of the cloud service user. For that reason, cloud computing operates on a shared responsibility model, where the provider manages certain aspects of security, while customers are responsible for others. Understanding and managing these responsibilities is crucial. It is very important to clearly define the division of among the cloud service provider and cloud service user. Providers typically secure the infrastructure, while customers have responsibility for their data and applications security in terms of compliance with security rules and proper use of security mechanisms. This understanding is essential for effective risk management.

- Network Security

  Cloud computing relies on network connectivity, making data vulnerable to interception during transit. This challenge is managed by cryptographic methods by implementation Virtual Private Clouds (VPCs), firewalls, and intrusion detection/prevention systems to secure network traffic. Use secure communication protocols and consider deploying a VPN for an additional layer of protection.

- Incident Response and Forensics

  Detecting and reacting to security incidents in a cloud environment require different approaches than traditional on-premises systems. Standard solution for this challenge is to develop and test incident response plans specifically tailored to the cloud environment. Cloud providers often offer tools for monitoring and logging that can aid in forensic analysis. Collaboration with the provider is crucial during incident response.

- Physical Security

Traditional data centers have physical security measures in place. In the cloud, the physical infrastructure is managed by the provider, raising concerns about the physical security of data. So, it is important to select reputable cloud providers that implement stringent physical security measures at their data centers and review and understand the provider's physical security certifications and practices.

– Data Loss Prevention
  Data Loss Prevention (DLP) considers prevention of unauthorized access, sharing, or loss of sensitive data in a cloud environment with multiple access points. Implementation of DLP solutions that monitor and control data movement within the cloud is one of the critical points of the system. This incloudes encryption, access controls, and real-time monitoring to identify and prevent data breaches.

– Continuous Monitoring and Auditing
  Ensuring the ongoing functionality and availability inclouding security of cloud resources requires continuous monitoring and auditing. Achievement of that goal assumes implementation of automated monitoring tools that provide real-time insights into the security posture of cloud resources. Regularly conduct audits and assessments to identify vulnerabilities and ensure adherence to security policies.

– Vendor Risk Management
  For the customer, relying on a third-party cloud provider introduces risks associated with the provider's security practices. For the customer it is important to conduct thorough due diligence when selecting a cloud provider. Evaluate their security controls, certifications, and compliance. Establish a robust vendor risk management program to monitor and manage ongoing security risks.

The relationship between information security and cloud computing is symbiotic. While cloud computing offers numerous benefits, organizations must proactively address the security challenges it introduces. A well-rounded security strategy, encompassing encryption, access controls, compliance management, and collaboration with cloud providers, is essential for mitigating risks and ensuring the confidentiality, integrity, and availability of data in the cloud. Regular assessment and adaptation to evolving security threats are integral components of a resilient security posture in a cloud-centric environment.

Chapter 3

Information security principles

Role of information security in the information systems and networks is safeguarding sensitive data, information systems and its resources from unauthorized access, alteration, and destruction. Cornerstone principles of information security are, [3]:

– Confidentiality

  Confidentiality ensures that information is accessible to those who have the right to access it and exclusively to them. Encryption techniques, access controls, and secure communication protocols are employed to prevent unauthorized access to sensitive data. This is particularly crucial for protecting personal, financial, and proprietary information. Almost all protection methods are based on cryptography

– Integrity

  Integrity ensures the accuracy and reliability of information by protecting it from unauthorized modification. Data validation checks, checksums, and hashing algorithms are used to verify the integrity of data. Version control mechanisms and access controls are implemented to prevent and detect unauthorized alterations. Beside other methods cryptography also has important role in this type of protection.

– Availability

  Availability ensures that information and information systems are accessible and usable by authorized users when needed. It is realized by redundancy, backups, and recovery plans in the event of disaster are

essential components to ensure continuity of operations. These measures mitigate the impact of hardware failures, natural disasters, and cyberattacks, ensuring information is consistently available.

- Authentication
  Authentication verifies the identity of users, systems, and devices to ensure that they are who or what they claim to be. It depends on the used user identification method and strong password policies, multi-factor authentication, and biometric identification methods are employed to validate user identities. This principle is fundamental for preventing unauthorized access.

- Authorization
  Authorization grants appropriate access to users approved based on their authenticated identity and role. Role-based access control (RBAC), based on the granting minimal possible level of privileges according to the defined role functionality, and regular access reviews are implemented to manage and restrict permissions. This guarantees that users can only access the resources essential for their designated role.

- Non-repudiation
  Non-repudiation ensures that a party cannot deny the authenticity of their digital signature or the origin of a message or transaction. Digital signatures, audit logs, and legal measures are utilized to establish accountability. This is crucial for legal and regulatory compliance, providing evidence of transactions and communications.

- Accountability
  Accountability traces and logs activities, assigning responsibility for actions taken within an information system. Robust logging, auditing, and monitoring systems are implemented to track user activities and system events. This principle helps in identifying and holding individuals accountable for security incidents.

- Security Awareness and Training
  Security awareness and training foster a security-conscious culture among users through education and regular awareness programs. Organizations conduct regular training sessions, simulate phishing attacks, and promote a security mindset among employees. Well-informed users are better equipped to detect and react to security threats.

– Security Governance

  Security governance establishes and maintains a framework that defines and enforces security policies, procedures, and controls across the organization. Organizations develop and enforce comprehensive security policies, conduct regular risk assessments, and establish a governance structure with clear roles and responsibilities. This ensures a systematic and consistent approach to security management.

– Incident Response and Management

  Incident response and management involve developing and implementing plans to respond to and recover from security incidents. Organizations establish dedicated incident response teams, define incident response procedures, and regularly test and update their incident response plans. This principle is crucial for minimizing the impact of security incidents and ensuring a timely and effective response.

These information security principles collectively offer a sturdy framework for organizations to establish and uphold a secure setting. Regular assessment, adaptation to emerging threats, and a commitment to continuous improvement are essential components of an effective information security strategy.

This comprehensive set of principles serves as a guide for organizations to establish and maintain a secure computing environment. Information security, based on the principles mentioned earlier, enables entities responsible for security within the system to determine who performed a particular action, when it was done, and what action was taken in the event of a security incident. Through forensic analysis, it becomes possible to identify the source of the problem and take appropriate measures to respond to the incident.

The confidentiality, integrity and availability are usually denoted al a CIA triade.

3.1 Cryptography and Information security

Cryptography and information security are intricately connected, playing pivotal roles in safeguarding digital communications, sensitive data, and overall cybersecurity. The relationship between cryptography and information security is multifaceted, encompassing various aspects such as confidentiality, integrity, authentication, and non-repudiation. Cryptography is one of

the many tools used in information security to achieve the goals of data and systems protection but cryptography is also an essential tool in information security. It provides a means to protect confidential data from illegal access. To achieve the goal of information security, it is important to use a layered approach that incloudes multiple tools and techniques. This ensures that even if one tool fails, there are other tools in place to protect the system/information.

3.1.1 Cryptography in brief

Since the advent of literacy and the transmission of messages written on different media, there has been a need to conceal the information that messages contain. The way the message is transformed so that the information contained in the message is available only to the person the message is intended for. For this purpose, various techniques were applied, starting from the transformation of the graphic representation of the message, the replacement of letters, to the imprinting of the message into the message carrier so that the message is imperceptible to uninitiated persons, invisible ink. The study and development of methods for covert communication is called cryptography meaning "secret writing" from the Greek words $\kappa\rho\iota\pi\tau o$ - meaning secret, concealed and $\gamma\rho\alpha\pi\eta\epsilon$ - meaning writing. In free translation, it denotes the art of secret writing. Cryptography as a human activity is very old and archaeological evidence of the transformation of written and pictorial messages dates back to 2000 BC, the era of ancient Egyptian civilization, [4], [5]. Over time, as the forms, means and scope of communication developed, the need of those for whom protected information was not intended for knowledge of protected content grew. This need has stimulated the development of new knowledge and skills in the field of secret communication, which is the analysis of protected messages and the development of methods for the reconstruction of protected information. Thus, in the field of secret communication, cryptanalysis, the twin sister of cryptography, was created. Their bond is unique and unbreakable because any progress in either area necessarily induced progress in the other. The interconnectedness of cryptography and cryptanalysis led to the formation of a unique area that deals with the study and development of secret communication on scientific principles called cryptology. The word cryptology is a compound word derived from the Greek word $\kappa\rho\upsilon\pi\tau o\varsigma$ which means secret, veiled, and the word "$\lambda o\gamma o\varsigma''$ which has the meaning of science, knowledge which loosely

translates to the science of secret communication.

Cryptography began its journey with the aim of protecting the information contained in communication messages, but the development of communication technologies has made the goals wider and the possibilities enriched. Modern cryptography in today's communication-networked world provides the following services:

– Confidentiality

– Data integrity

– Authentication

– Non-repudiation

Cryptology studies and constructs message transformations whose application enables stated goals.

Cryptography defines message transformations as follows.

Let us denote by $A = \{a_1, a_2, ..., a_l\}$ an arbitrary alphabet whose symbols are$a_1, a_2, ..., a_l$.

Then we denote the set words whose length is i made from the letters in the alphabet A with $A^i = \{x_1x_2 \dots x_i \,|\, x_j \in A,\ j = 1, 2, ..., i\}$ and with

$$A^* = \bigcup_{i=0}^{\infty} A^i$$

the set of all words made from the letters in the alphabet A.

Let the three alphabets be given, M alphabet of plaintext messages, K alphabet of keys and C alphabet of ciphetexts. Then any mapping F such that $F : K^* \times M^* \longrightarrow K^* \times C^*$ and $F(k, m) = (k, c)$ is called the transformation of the message $m \in M$ by applying the function F and the key $k \in K$. In literature, it is common for the key k to be considered as a parameter and this is indicated by $F_k(m) = c$.

In order to realize the functionalities of CIA and undeniability of transformation, they must possess certain properties. For example, to realize privacy for a given function F, which we call the encryption function, there must be a function F^{-1} , which we call the decryption function, and for which it applies

$$F^{-1}(k_2, F(k_1, m)) = (k_1, m).$$

The encryption function is usually denoted with E and the decryption function with D, so the previous equality can also be written as follows

$$D_{k_2}(E_{k_1}(m)) = (k_1, m)$$

Encryption and decryption transformations are commonly referred to as cryptographic algorithms.

3.1.2 Classification of cryptographic algorithms

The classification of cryptographic algorithms is carried out in two aspects, equality of the applied cryptographic keys and level of security achieved by the algorithm.

In the relation of equality of cryptographic keys k_1, k_2 used for encryption and decryption cryptographic algorithms are classified as:

- The class of cryptographic algorithms in which the keys k_1 and k_2 are equal are called the class of symmetric cryptographic algorithms ($k_1 = k_2$).
- The class of cryptographic algorithms where the keys k_1 and k_2 are different are called the class of asymmetric cryptographic algorithms ($k_1 \neq k_2$).

Regarding the achieved security level cryptographic algorithms are classified as:

- Perfectly secure cryptographic algorithms, the algorithms which cannot be broken and information disclosed never even in the case that attacker has unbounded computational power, Information theoretic security.
- Computational secure cryptographic algorithms which theoretically can be broken but it is impossible to concentrate necessary resources (time, memory, processor power) - Computational security.

Process of security level evalaution of cryptographic algorithm is serious research task. Starting point in that process is famous Kerckhoffs's principle who claims that cryptographic algorithm must be secure even in the case when all his elements are known to the attacker except the applied cryptographic key. From that approach importance of cryptographic keys and methods for their generation stems.

3.1.3 Cryptographic algorithm parameters and continuous auditing

In the context of information security systems auditing generation and establishment of cryptographic keys/security parameters significantly increases price and necessary time. External auditing also assumes disclosure of large amount of sensitive data to auditors as a trusted third party.

By outsourcing routine audit tasks to third parties, internal resources can be used more strategically and beneficial to the organization. But, outsourcing Information security audit and automatization of it imply proven security methods usage, auditable and secure methods of their implementation.

From an information security standpoint, a whole range of security challenges arise, starting with security goals and security architecture through their operationalization and implementation. This is particularly reflective of the information security audit as part of the audit of information systems. In terms of information security cryptographic algorithms and cryptographic protocols are significantly standardized and support the approach of continuous external audit and improvement of the security of the subject information system. On the other hand, all of these solutions involve the use of cryptographic parameters created appropriately and under certain conditions. This audit segment requires specialist knowledge and the ability to assess the adequacy of the procedures applied. Contrary to cryptographic algorithms and protocols in this segment, there is no generally accepted standardization. This research is an attempt to develop a method that would be reliable in theoretical terms and proofs and also independent of trusted third parties. Such a method would significantly improve the possibilities of continuous revision in this segment and information security in the systematic sense.

Chapter 4

Information theory approach to cryptographic parameters generation

As we have previously mentioned if theKerckhoffs' principle is taken into account, cryptographic parameters with their quality represent, along with the default quality of cryptographic algorithms, guarantees the reliability of implemented security solutions. In the light of managing cryptographic parameters, the procedures of generating them are a challenge that is constantly being posed to the designers of security solutions. The development of technology and scientific knowledge has also initiated the development of techniques for generating and distributing cryptographic parameters. Traditional cryptology, until the advent of asymmetric cryptographic algorithms, knew three models for establishing cryptographic keys:

- Face to face exchange of cryptographic keys when participants of communication, mark them with A and B, exchange keys in direct contact, Figure 4.1.

- A centralized distribution of keys in which at the request of one communication participant, say A asks for a key to communicate with B , a dedicated trusted party for the distribution of keys, KDC , generates a key and distributes it up to A and B in a reliable way, Figure 4.2.

Figure 4.1: Cryptographic key exchange in direct contact

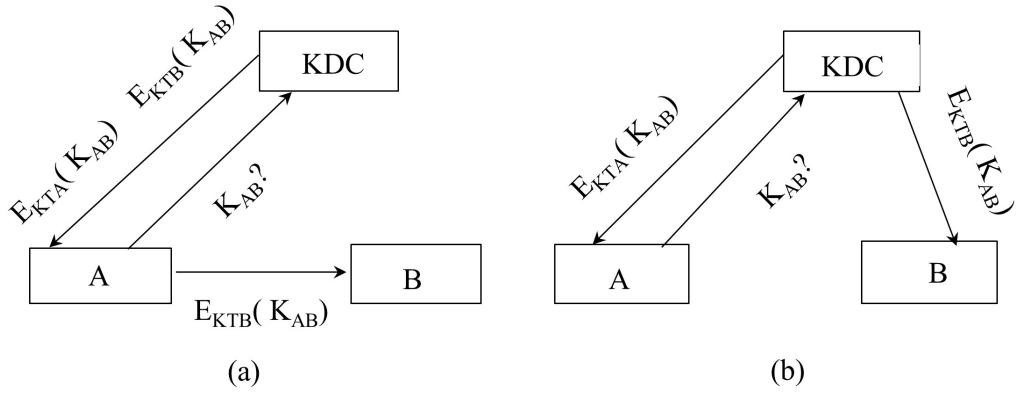

Figure 4.2: Centralized exchange of cryptographic keys

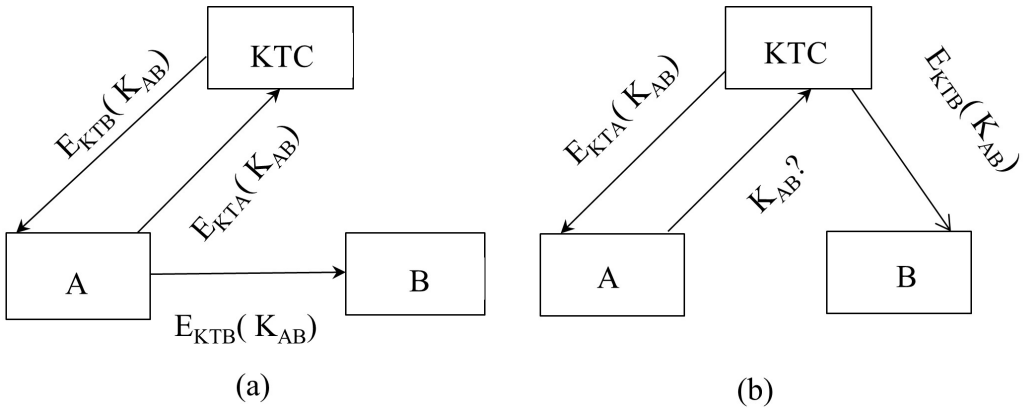

Figure 4.3: Key translation method for cryptographic key exchange

- A dedicated communication network in which A to communicate with B in one scenario A create cryptographic key K_{AB} and KTC translate it to in secure fashion B, or in another scenario A create cryptographic key K_{AB}, transmit it to KTC who encrypt it with B cryptograpfic key and that encrypted message send to A to reliably forward it to B , Figure 4.3.

The development of information and communication technologies and the rapid increase in the number of participants in communication networks have induced a number of challenges in terms of communication security and confidentiality of data contained therein. A series of difficulties in process of cryptographic keys creation and its delivery to users are immanent to the classical model has surfaced with modern communication networks, [5]:

- Initialization of the system involves the assignment of cryptographic keys to all participants in the system in a reliable and secure way. In the context of a classical distribution, this implies either mass direct contacts or the existence of a dedicated commonly trusted entity for key exchange and the primary initialization of such a system, which in turn poses a challenge from the point of view of primary key distribution. The usual and the simplest approach to solving this challenge is the application of courier delivery, but in mass networks, such an approach is not practically applicable. A graphical representation of the direct exchange of cryptographic keys is given in the Figure 4.1

- The existence of a centralized infrastructure for producing and disseminating cryptographic keys implies use of a single entity as a commonly trusted party, which in the world of shared interests (business, politics,...) is not easy to achieve or realistically expect. A graphical representation of the centralized distribution of cryptographic keys is given in Figure 4.2

- The creation of solutions that mitigate induced challenges has led to the creation of complex mixed models of distribution of cryptographic keys that have a rather complex organizational structure and demanding administration and maintenance procedures. That management complexity mitigate widespread and massive implementation A graphical representation of the centralized distribution of cryptographic keys is given in the Figure 4.3

The aforementioned prompted the search for new ideas, which resulted in the discovery of new protocols for the establishment and exchange of cryptographic keys.

With the advent of asymmetric cryptographic algorithms, a technique called the ''Digital Envelope" came into play, [6]. Since asymmetric cryptographic algorithms are, by security criterion, computationally secure cryptographic algorithms, this technique did not quite meet the requirements for the generated keys to be completely secured which is a requirement in the process of evaluating security solutions.

It has been shown, in a theoretical and partially limited practical sense, that such solutions are possible and an example of this is the algorithm for the quantum distribution of cryptographic keys. This fact has led the research community to analyze the possibilities for creating practically applicable protocols for establishing cryptographic keys that can be mathematically proven to be absolutely safe for their users, that is, that no one other than the executors is able to possess any information about generated key contents. Analyses and attempts are directed by the following requirements:

- The communication takes place exclusively between potential users of the requested cryptographic key, A and B according to the previous labels, without the presence and influence of a commonly trusted party, thus enabling the realization of cryptographic traffic protection from end to end of the communication line.

- Communication takes place through the use of public, unprotected communication channels so that the potential attacker is able to see all the exchanged messages, but by its nature in relation to the protocol is passive, which means that there is no possibility to insert or exchange messages without legitimate participants noticing it, a publically available authenticated channel.

- The protocols that are the subject of analysis can be represented by mathematical models that allow the derivation of formal claims about the degree of their security. The preferred theoretical framework of this modeling is Information theory.

4.1 Cryptographic key establishment protocols based on Information theory overview

For the first time in the history of cryptography, Claude Shannon defined the security of cryptographic algorithms in his works in an exact way, [7] using Probability and Information theory. The beauty of Shannon's idea and its revolutionary nature lies in the fact that the power of the cryptographic algorithm is defined independently of the resources which the attackers possess, [8], [9]. This means that if the cryptographic algorithm is secure in Shannon's sense then it is resistant to all possible types of attacks and we refered to it as Information theory Security Model (ITSM). Shannon's ideas were one of the guidelines in thoughts on formulating protocols for establishing cryptographic keys.

But the situation with the protocols for establishing cryptographic keys is not exactly the same as with cryptographic algorithms, because the protocol takes place through a publically available authenticated channel. The idea in this new context is that the amount of information that an illegitimate user can obtain by analyzing messages collected from a publically available authenticated channel during the execution of the protocol is not enough to compromise the cryptographic key agreed between legitimate users in an effective way. The imperative for technological security in cyberspace has led to the acknowledgment of the significance of protocols for establishing cryptographic keys, thereby intensifying research efforts in this domain.

The first papers in this direction were published by Wyner, [10] and later refined by the Maurer, [11] and Ahlswede, [12], defining a formal model for

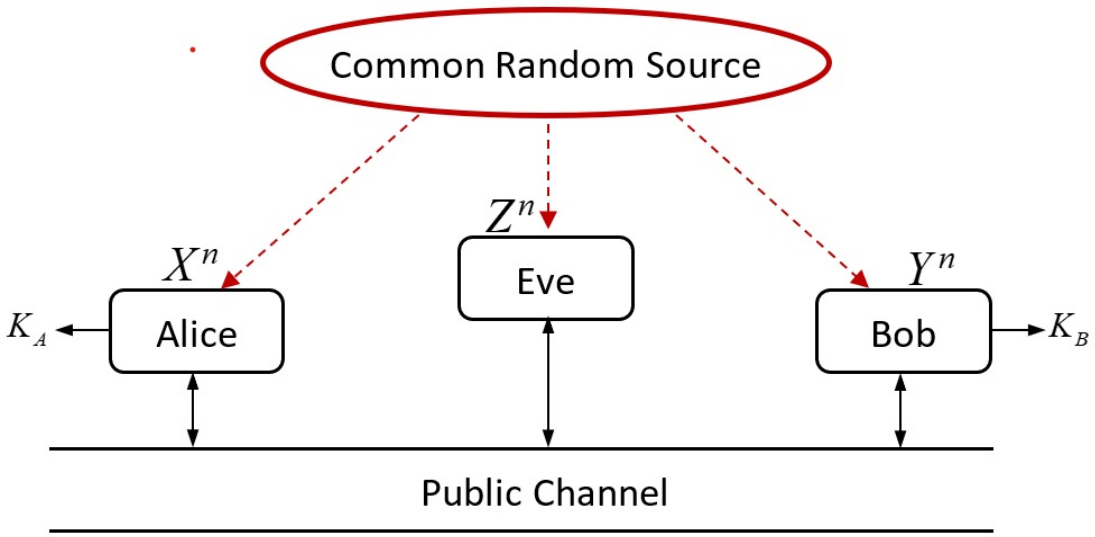

Figure 4.4: Maurer Satellite model for reliable key establishment

this type of protocol. As in the cryptographic literature, legitimate protocol actors who want to establish a cryptographic key are called Alice and Bob and are labeled with A, B respectively. An illegitimate curious observer of the protocol is called Eve and denotes with E. Eve wants to find out the cryptographic key established by the protocol between Alice and Bob so that she can track and know what information they exchange in their later communication because this is the only way in case the cryptographic algorithm is reliable.

The communication model described in [11] is shown in the Figure 4.4.

The input data for the execution of the cryptographic key establishment protocol Alice, Bob and Eve are obtained from a common source of coincidence (CSR) U. These are sequences of symbols of length n, let's mark them in order with $X^n = \{x_1, x_2, \ldots, x_n\}$, $Y^n = \{y_1, y_2, \ldots, y_n\}, Z^n = \{z_1, z_2, \ldots, z_n\}$, which they belong to Alice, Bob and Eve respectively.

When the protocol is executed, a string of binary symbols K_A, K_B with the following properties are obtained:

- Arrays K_A, K_B are equal to probability close to the unit i.e. $P(K_A = K_B) \approx 1$. It is possible to define a procedure for checking the equality of the resulting strings of symbols and in the event that the arrays are not identical, additionally carry out a process by which a series of identical

symbols is obtained on both sides, which indicates $K^{m(n)}$ $m\,(n)$ denotes the matching symbols array length both sides.

- The result of the protocol execution, the cryptographic key $K^{m(n)}$, is secure, which implies that Eve as an observer does not have enough information to compromise the derived key formally expressed by

$$I\left(K^{m(n)}, Z^n\right) = 0$$

From the point of view of usability, this condition is quite strict and in practice its weakened variant is used

$$\lim_{n\to\infty} I\left(K^{m(n)}, Z^n\right) = 0$$

assuming that n is the length of binary symbol sequence. The meaning of this equation is that Eve, possibly, possesses some amount of information but it is not enough to compromise the derived key.

For a detailed explanation of this topic, see [9].

4.2 Protocol architecture for establishing cryptographic keys in an Information theory model

This type of protocol is executed in several stages.

The first stage involves the creation of a series of symbols of length $\not{n}$ by a CRS U that we denote as $U^n = \{u_1, u_2, \ldots, u_n\}$ a sequence of independent realizations of a random variable U. The generated sequence U^n is passed via three independent BSM to Alice, Bob and Eve, which we refer to with $X^n = \{x_1, x_2, \ldots, x_n\}$, $Y^n = \{y_1, y_2, \ldots, y_n\}$, $Z^n = \{z_1, z_2, \ldots, z_n\}$. By forwarding an array U^n, errors can occur on the communication channels and accordingly the strings X^n, Y^n, Z^n do not have to be mutually identical nor with $.U^n$ The statistics of the communication channel is given with (P_{XZY}, X, Y, Z) .

The second phase is aimed at increasing the similarity among Alice's and Bob's sequence. This phase is realized through messaging between Alice and Bob using a publically available authenticated channel whereby Eve knows

all the exchanged messages and actions that legitimate participants apply but not the initial content of their strings. At the same time, despite knowing the steps of protocol and the freedom of action, the similarity of the Eve sequence with the Alice/Bob sequence decreases or at least do not increase. A measure of similarity is used as a measure of the distance between strings of symbols. In practice, this is most often Heming's distance. In the literature procedure is named as Advantage distillation.

Next phase of the protocol, third one, aims to recognize and extract identical parts in sequences created after the second phase at Alice and Bob. This process takes place by exchanging messages with each other through a publically available authenticated channel, provided that the amount of information Eve extracts while monitoring that communication does not grow or remains small enough. At the end of this phase, which is called Information reconciliation, Alice and Bob have sequences that are identical with high probability.

The fourth, final phase of the protocol, called Security amplification, consists in performing a common key on both sides by applying a publically known and pre-agreed procedure.

We will describe in more detail each phase.

4.2.1 Common randomness sequence generation

Given Kerckhoffs's principle, the basis for message security, along with the default high-quality cryptographic algorithm, is that cryptographic keys are dimensioned and generated to have maximum entropy that prevents attackers from reconstructing it in any effective way. Provided that the key length is greater than or equal to the length of the message, systems are obtained that are known to be absolutely safe. As we mentioned earlier, this incloudes attackers with unlimited computing power and those who have access to quantum computers, [Grover]. Within probability theory and mathematical statistics, techniques have been developed to transform the realization of random processes with non-uniform probability distribution into the realization of random variables with uniform probability distribution. Using those procedures it is possible to obtain samples with uniform distribution of probabilities from non-uniform ones. In Maurer's satellite model, the assumption is that the realizations of a random process are U values from the set $\{0,1\}$, of essences, and that probability $P\left(U=0\right)=P\left(U=0\right)=\frac{1}{2}$. As a source of chance U , different processes can be used and academic literature classifies

them into two types:

1. Sources that extract coincidence from processes not related to the communication channel - the source model, such as in [13], [14]

2. Sources that extract coincidence from processes related to a transmission channel - a channel model such as in [10].

4.2.2 Legitimate participants mutual information increasing protocol

According to Maurer's satellite model, image 4.4, we imply that a series of symbols U^n to Alice, Bob and Eve is transmitted over three independent BSCs defined by the probabilities of channel error p_A, p_B, p_E respectively whereby the $0 < p_A, p_B, p_E < \frac{1}{2}$. Sequences obtained by Alice, Bob and Eve we denote with X^n, Y^n, Z^n. The consequence of this approach is the fact that the sequence Y^n can be considered the result of the transmission of X^n the array over a BSC characterized by the probability of error

$$p_{AB} = p_B \cdot (1 - p_A) + (1 - p_B) \cdot p_A$$

Namely

$$\begin{aligned} & p_{AB} = P\left(x_i \neq y_i\right) = \\ & = P\left(x_i \neq y_i \,|x_i = u_i\right) \cdot P\left(x_i = u_i\right) + P\left(x_i \neq y_i \,|x_i \neq u_i\right) \cdot P\left(x_i \neq u_i\right) \\ & = P\left(y_i \neq u_i\right) \cdot P\left(x_i = u_i\right) + P\left(y_i = u_i\right) \cdot P\left(x_i \neq u_i\right) = \\ & = p_B \cdot (1 - p_A) + (1 - p_B) \cdot p_A \end{aligned} \tag{4.1}$$

what the position is explained.
The relationship between Alice and Eve can be treated in the same way, Z^n it can be considered the result of the transfer of the array X^n over a BSC characterized by the probability of error

$$p_{AZ} = p_Z \cdot (1 - p_A) + (1 - p_Z) \cdot p_A.$$

In such a defined environment we are interested in the situation when the $0 < p_E < min\left\{p_A, p_B\right\}$. Reason for this is the fact that when $\max\left\{p_A, p_B\right\} < p_E$ in the paper [10] it is shown that it is possible to achieve confidential

communication and the situation when it $min\{p_A, p_B\} < p_E < \max\{p_A, p_B\}$ can be considered better for Alice and Bob than the situation when $0 < p_E < min\{p_A, p_B\}$ a positive result in this case allows confidential communication in both cases. It should be noted here that during this phase, comparison and equalization with the sequence U^n are not carried out, but sequences X^n, Y^n with each other.

The idea of this part that Alice and Bob, communicating through a publically available authenticated channel in their strings X^n, Y^n , detect segments of sequences of symbols where the probability of differentiating in the same positions in strings X^n, Y^n will be less than (4.1) and that, on the other hand, Eve is unable to gain enough information from their communication to make the probability of differentiation X^n, Z^n in positions held by Alice and Bob is less than their error.

More protocols have been defined for this purpose and we will describe some of them below.

The repetition code advantage distillation protocol (RCAD)

In the literature, this protocol is encountered in different, mutually equivalent, forms [11], [15], [16]. This protocol proceeds as follows:

1. Alice picks the number N so that $N = k \cdot N$ and she tell picked value to Bob through a publically available authenticated channel.

2. Alice divides X^n, her series into k equal parts and processes each in a next way:

   a) For the currently processed segment of the , Alice pick up randomly a bit q $\left(P\left(q=0\right) = P\left(q=1\right) = \frac{1}{2}\right)$ and a code word $Q^N = \left(\overbrace{q, q, \ldots q}^{N}\right)$.
   Then calculate

   $$X^N + Q^N = (q + x_1, q + x_2, \ldots, q + x_N)$$

   The calculated string is sent to Bob.

b) After receiving string calculated by Alice, Bob computes

$$Y^N + X^N + Q^N = (y_1 + q + x_1, y_2 + q + x_2, \ldots, y_N + q + x_N)$$

If as a result Bob gets $\left(\overbrace{q, q, ...q}^{N}\right)$ where $q \in \{0, 1\}$ he thinks his and Alice's streak match. As a feedback to Alice Bob forwards the bit F determined by

$$F = \begin{cases} 1 & \text{computed vector is in the form } \left(\overbrace{q, q, ...q}^{N}\right) \\ 0 & \text{otherwise} \end{cases} .$$

3. If $F = 1$ that part of the array is accepted as equal on both sides, Alice and Bob, continue to participate in creation of bit series for the next iteration, otherwise that string of bits is rejected from the next process iteration.

4. The protocol is repeated until all segments are processed.

The parameter N is selected so that the length of the sample n is its product and additional criteria are maximization of the likelihood probability and minimization of the amount of information flowing towards Eve. As is the logical, probability depends on the characterization of communication channels and in our case it can be shown that [15], [17]:

– The probability of accepting length segments N that differ is, the probability of error,

$$p_{AB}^{RCAD} = \frac{(p_{AB})^N}{(p_{AB})^N + (1 - p_{AB})^N} \tag{4.2}$$

where p_{AB} the probability of error is a distinction in one position in the segment.

– Because of the way Alice and Bob communicate, every transmitted segment of the length N is known to Eve and she can extract some information from it. By calculating the value of the expression

$$Z^N + X^N + R^N$$

Eve gets a segment that can pair with her array and approximate the results that Alice and Bob get. The probability of distinguishing between Alice and Eve's segments is given by:

$$p_{AE}^{RCAD} = \frac{1}{(p_{AB})^N + (1-p_{AB})^N} \cdot \sum_{w=\left\lceil \frac{N}{2} \right\rceil}^{N} \left(\begin{array}{c} N \\ w \end{array} \right) p_w$$

where the p_w is probability that the string of length N has weight w.

– The effectiveness of this procedure can be expressed through the ratio of the length of the starting sequence to the length of resulting sequence upon the procedure execution and it is shown that the coefficient of efficiency is given with, [17]

$$\mu_{AB}^{RCAD}(p_{AB}) = \frac{(p_{AB})^N + (1-p_{AB})^N}{N}.$$

The importance of this protocol is reflected in the first place in its existence, that a protocol can be defined and meet its goals in the specified environment, the growth of the similarity of the sequences of Alice and Bob, and the increase of the difference with Eve's sequence. In this way, it has been proven that private communication through a publically available authenticated channel is possible, which results in the possibility of establishing cryptographic keys through a publically available authenticated channel and eliminating trusted third parties from the process. From the point of view of the effectiveness and applicability of this protocol, it must be noted that its effectiveness is not great and therefore has theoretical significance.

The bit pair iteration advantage distillation protocol (BPIAD)

This protocol is by its very nature iterative. The input strings into $s-$th iteration will be marked with X^{n_s}, Y^{n_s}. Accordingly, it is in the first iteration $X^{n_1} = X^n, Y^{n_1} = Y^n$ and the output from the i-th iteration is the entrance to $i+1$ the iteration, except in the case of the last iteration. The final iteration

gives as a result of this phase sequences of Alice and Bob that are significantly different from Eve's sequence. The protocol proceeds as follows, [17]:

1. X^{n_s}, Y^{n_s} there are input strings for $s-$ th iteration. Alice and Bob devide their sequences in chunks of every two adjacent bits, the chunks are disjointed. For each chunk, Alice calculates its *xor* value, forms a set of bits thus obtained

$$\left\{ X_{2i+1}^{n_s} \oplus X_{2i}^{n_s}, i = 0, 1, \ldots, \left\lfloor \frac{n_s}{2} \right\rfloor \right\}$$

   He sent it to Bob.

2. Bob performs the same operation with the essences of his sequence, gets a set

$$\left\{ Y_{2i+1}^{n_s} \oplus Y_{2i}^{n_s}, i = 0, 1, \ldots, \left\lfloor \frac{n_s}{2} \right\rfloor \right\}$$

   and compares them to Alicia's. The selection rule is defined by

   – Let it be $i = 0, 1, \ldots, \left\lfloor \frac{n_s}{2} \right\rfloor$. if the $X_{2i+1}^{n_s} \oplus X_{2i}^{n_s} \neq Y_{2i+1}^{n_s} \oplus Y_{2i}^{n_s}$ chunks $X_{2i+1}^{n_s}, X_{2i}^{n_s}$ and $Y_{2i+1}^{n_s}, Y_{2i}^{n_s}$ are both rejected and do not participate in the further iterations

   – Let it be $i = 0, 1, \ldots, \left\lfloor \frac{n_s}{2} \right\rfloor$ if the $X_{2i+1}^{n_s} \oplus X_{2i}^{n_s} = Y_{2i+1}^{n_s} \oplus Y_{2i}^{n_s}$ bit $X_{2i+1}^{n_s}$ is take in Alice's sequence and the bit $Y_{2i+1}^{n_s}$ is take in Bob's sequence for the next iteration $s + 1$.

It can be shown that at the end of $s-$ th iteration next statements are valid, [15], [17]:

– The probability of an error in the same positions after $s-$ th iteration in the arrays X^{n_s+1} and Y^{n_s+1} is given with

$$p_{AB_s}^{BPIAD} = \frac{\left(p_{AB_0}\right)^{2^s}}{\left(p_{AB_0}\right)^{2^s} + \left(1 - p_{AB_0}\right)^{2^s}} \tag{4.3}$$

– The probability of an error in the same positions after $s-$ that iteration in the arrays X^{n_s+1} and Z^{n_s+1} is given with

$$p_{AE_s}^{BPIAD} = p_{AE_0}^{BPIAD} \tag{4.4}$$

– The effectiveness of the data protocol is

$$\mu_{AB_s}^{RCAD}\left(p_{AB_0}\right) = \frac{\left(p_{AB_0}\right)^2 + \left(1 - p_{AB_0}\right)^2}{2^s} \tag{4.5}$$

Equality (4.3) with the analysis of the above protocol shows that by analyzing the described protocol and the results mentioned, we see that $p_{AB_s}^{BPIAD}$ depends on the size of the initial error of Alice and Bob's sequence p_{AB_0}. The greater the initial error requires a larger number of iterations to achieve an information supremacy over Eve, $p_{AB_s}^{BPIAD} < p_{AE_s}^{BPIAD}$. Equality (4.5) shows that the length of the resulting sequence after each iteration reduces exponentially by factor two.

The bit pair iteration advantage distillation/degeneration protocol (BPIADD)

This protocol is also of an iterative type. As we have stated, equality (4.4), shows that the amount of information Eva has during the execution of this protocol does not decrease compared to the amount of information it possesses at the beginning. The question arises as to whether it is possible to define a protocol during which the amount of information that Eve possesses will be reduced with each iteration of the execution of the protocol. If this were possible, then Alice and Bob would be in a much more favorable position at the time of launching the second phase of the cryptographic key establishment process, Information Reconciliation.

It is shown that this is possible and the solution is the BPIADD protocol defined in the paper [wang2015].

As in the BPIAD protocol definition, the input arrays into $s-$ th iteration will be marked with X^{n_s}, Y^{n_s}. Accordingly, it is in the first iteration $X^{n_1} = X^n, Y^{n_1} = Y^n$ and the output from the i -th iteration is the input to the iteration $i+1$, except in the case of the last iteration. The final iteration gives the result of this phase, sequences of Alice and Bob that differ significantly from Eve's sequence. The protocol proceeds as follows, [17]:

1. X^{n_s}, Y^{n_s} there are input strings for $s-$th iteration. Alice and Bob form blocks of every two adjacent bits, the blocks are disjointed.

2. For each block, Alice calculates its *xor* values, forms a set of bits thus obtained $\left\{A_i = X_{2i+1}^{n_s} \oplus X_{2i}^{n_s}, i = 0, 1, \ldots, \left\lfloor \frac{n_s}{2} \right\rfloor\right\}$ and sends it to Bob.

3. Bob carries out the same operation with the values of his sequence and forms $\left\{B_i = Y_{2i+1}^{n_s} \oplus Y_{2i}^{n_s}, i = 0, 1, \ldots, \left\lfloor \frac{n_s}{2} \right\rfloor\right\}$ and sends it to Alice.

4. For the $i = 1, 2, \ldots$ following procedure is carried out for everyone:

   – If $A_i \neq B_i$ then Alice deletes $X_{2i+1}^{n_s}, X_{2i}^{n_s}$ from X^{n_s} and Bob wipes $Y_{2i+1}^{n_s}, Y_{2i}^{n_s}$ out of Y^{n_s}.

   – If then $A_i = B_i$ Alice checks if she is $X_{2i}^{n_s} = 1$. If she is then delete $X_{2i}^{n_s}$ from X^{n_s} and if she is not delete $X_{2i+1}^{n_s}$ from X^{n_s}. Bob performs the same action with his string Y^{n_s}.

It appears that following the initial iteration of this protocol, the subsequent relationships hold true, [17]:

– After the first iteration of the protocol, the probability of an error between Alice and Bob's sequence at $i-$ th position is given by

$$p_{AB_1}^{BPIADD} = \frac{1}{2} \cdot \frac{\left(p_{AB_0}\right)^2}{\left(p_{AB_0}\right)^2 + \left(1 - p_{AB_0}\right)^2} \tag{4.6}$$

– After the first iteration of the protocol, the probability of an error between Alice and Eve's sequence at $i-$ th position is given by

$$p_{AE_1}^{BPIADD} = \frac{p_{AE_0}}{2} + p_{AE_0}\left(1 - p_{AE_0}\right) > p_{AE_0} \tag{4.7}$$

– After the first iteration of the protocol, the coefficient of efficiency is

$$\mu_{AB_1}^{RCAD}\left(p_{AB_0}\right) = \frac{\left(p_{AB_0}\right)^2 + \left(1 - p_{AB_0}\right)^2}{2} \tag{4.8}$$

From the expression (4.6) it is easy to see that $p_{AB_1}^{BPIADD} < p_{AB_0}$ that is, that the probability of a distinction in the Alice and Bob sequences decreases in $i-$ th place and from the expression (4.7) that $p_{AE_1}^{BPIADD} > p_{AE_0}$ the probability of a distinction in the Alice and Eve sequence increases there.

In the following iterations using (4.6)–(4.8), it follows that in $s-$ th iteration it is in order

$$\begin{aligned} p_{AB_s}^{BPIADD} &= \frac{1}{2} \cdot \frac{\left(p_{AB_{s-1}}\right)^2}{\left(p_{AB_{s-1}}\right)^2 + \left(1 - p_{AB_{s-1}}\right)^2} < p_{AB_{s-1}} \\ p_{AE_s}^{BPIADD} &= \frac{p_{AE_{s-1}}}{2} + p_{AE_{s-1}}\left(1 - p_{AE_{s-1}}\right) > p_{AE_{s-1}} \\ \mu_{AB_s}^{RCAD}\left(p_{AB_{s-1}}\right) &= \frac{\left(p_{AB_{s-1}}\right)^2 + \left(1 - p_{AB_{s-1}}\right)^2}{2} \end{aligned} \tag{4.9}$$

Relations in the (4.9) shows that the probability of error between Eve and Alice increases relatively quickly and the probability of error between Alice and Bob's sequences reduces relatively quickly and the amount of information about the sequences entering the next stage available to Eve does not affect the derived cryptographic key. Also, the feature of this protocol is that it favors units so that the number of iterations must be carefully selected so as not to disturb the desired probability distribution of zeros and ones.

Information reconciliation

With the completion of phase two, Advantage distillation, Alice and Bob created sequences so that Bob has more information about Alice's sequence than Eve has. This phase of the protocol aims to detect and eliminate any existing differences in the sequences of bits that participate in this phase of cryptographic key alignment by exchanging messages through a publically available authenticated channel. And at this stage, it goes without saying that the messages between Alice and Bob are fully accessible to Eve. The amount of information that Eve can extract from this communication about the sequences with which Alice and Bob operate directly affects the length of the strings that will be used to create the final cryptographic key. The length of the string from which the cryptographic key will be constructed is inversely proportional to the amount of information Eve possesses.

After the publication of the first paper describing such a protocol, [18], protocols based on different ideas appeared in the academic literature. In

practice, those protocols that offered efficiency in their execution came to life. We will briefly describe a few of the most widespread in practice:

– In this context, the most widespread use is of the CASCADE protocol described in the paper [19]. The protocol takes place in multiple steps and is iterative by its nature. Let X^{n_s}, Y^{n_s} the strings possessed by Alice and Bob at the beginning of this protocol. Also, Alice and Bob possess a pre-arranged and publically known permutation Π of the required length. In $i-$th iteration, the protocol proceeds as follows:

  1. Alice and Bob arrays $X_i^{n_s}, Y_i^{n_s}$, each their own, are permutated by permutation Π and receive sequences $\Pi\left(X_i^{n_s}\right), \Pi\left(Y_i^{n_s}\right)$. On permuted sequences Alice and Bob, again each their own, divided into blocks of bit length k_i , $\left(x_1^j, x_2^j, \ldots, x_{k_i}^j\right), \left(y_1^j, y_2^j, \ldots, y_{k_i}^j\right)$ respectively. Now Alice for her blocks of calculations *xor* value,

     $$x_1^j \oplus x_2^j \oplus \cdots \oplus x_{k_i}^j = PX^j$$

     Pass it on to Bob. Bob also calculated his *xor* bits,

     $$y_1^j \oplus y_2^j \oplus \cdots \oplus y_{k_i}^j = PY^j$$

  2. Bob checks the equality of the computed values for his and Alice's blocks and if at some position the pairs are not equal, $PX^j \neq PY^j$, Alice and Bob using a binary search algorithm, divide the block into sub blocks, recalculating the parity values under the newly obtained blocks and exchanging that data in order to find a bit in the block that differs. Upon error bit detection Bob change its value.

  Now there's a cascading effect on the scene. Namely the altered bit was involved in the previous iteration and the change in its value in $i-$th iteration affects Bob's parity bit in the previous, $i-1$ iteration. Bob performs a recalculation of the mate bits from the previous iteration and using the already obtained mate bits from Alice, with possibly some additional communication, reveals some more errors if they exist and changes the value of these bits. This process of backwards is called the cascading effect. The described procedure is executed for each block in which there is a disagreement to pair bits. Processing all

blocks in which there iare a disagreement to pair bits ends this iteration and moves into iteration $i+1$. For iteration $i+1$ is taken $k_{i+1} = 2 \cdot k_i$ and the procedure is repeated. The block k_i, length parameter, in this process, fundamentally affects the effectiveness of this protocol. Numerous theoretical and experimental considerations have addressed the problem of the effectiveness of this protocol with the aim of achieving the most efficient procedure, [20], [21], [22], [23]. Sigumoto defined the modified procedure, [24], [25], and obtained results showing that his procedure was very close to the theoretical limit of the effectiveness of this procedure. The results of his analyses show that four iterations of this protocol are sufficient to achieve equality between Alice and Bob series. The negative feature of this procedure is reflected in the great communication complexity that has consequences of twofold in terms of efficiency, duration time and reduction of the string length for key execution in order to make available as little information as possible to Eve.

– As communication complexity is a fundamental weakness of the CASCADE protocol, both in terms of runtime and in the key material length, different approaches have been attempted to overcome this problem. One of the solutions appeared as the Winnow protocol described in [26]. The idea of this protocol is to replace the binary search algorithm in the process of detecting and correcting the wrong bit with Heming's error detection and correction code, [27]. The process proceeds as follows.

  1. The strings of bits X^n, Y^n they possess, Alice and Bob divide into blocks whose length are equal to the length of the Heming error-correcting code word with the generator matrices G and H connected by the
  $$G \cdot H^T = 0$$
  designating generator and error-checking matrix respectively..
  2. Let's mark M_a M_b Alice and Bob's corresponding block respectively and the corresponding syndromes with S_a and S_b , computed using the matrces G and H. Alice passes to Bob S_a and he computes now by gaining that data $S_d = S_a \oplus S_b$.
  3. If they are $S_d = 0$ M_a M_b considered equal

4. If, on the other hand $S_d \neq 0$, Bob makes the minimum number of changes in his block, M_b, gets the modified block M_b' and for the newly obtained block recalculates the syndrome S_b until he gets $S_d = 0$.

Conducted analyses regarding protocol characteristics, execution speed, and realized string length for performing cryptographic key amounts of information flowing towards Eve have shown that the application of this technique to detect and correct errors produces good results, [28]

– Research into the problem of performing cryptographic keys at first focused mainly on the possibility of finding and realizing solutions as such. They were less concerned on the physical attributes of the environment where protocols of this kind are implemented. Over the past few years, protocols of this type have been considered that are executed in environments where there are significant limitations in terms of packet loss during transmission packets, constrained protocol time execution and constrained communication and computational complexity. For satellite communication, internet of things, sensor communication networks. Using the ideas used in the solution described above and in the context of restrictive conditions, Gallagher's Low Density Parity Check codes, [29], was imposed as a possible solution. This idea was first used in [30]. The comparative superiority of LDPC codes in restrictive environments is that they require extremely little communication complexity and can also be realized in conditions of significant disproportion of the computational capabilities of communication trays. On the other hand, decoding of LDPC codes takes place with the involvement of higher requirements of process and memory resources compared to the CASCADE and Winnow protocols but this compensates for a significant reduction in communication complexity and an increase in sequence length for performing a cryptographic key because reduced communication complexity affects the amount of information flowing to Eve.
Detailed information regarding this approach can be found in [31].

One of the ideas that is significantly exploited in this context is the application of neural networks to detect and correct errors and its application can be found in [Mehic][mehic28].

This issue with a detailed overview of ideas and a list of references was addressed in , [32], [28], [33].

Privacy amplification

The Privacy amplification protocol is the final step in the procedure for establishing the key between Alice and Bob.

Starting from the sequences X^n, Y^n Alice and Bob after a series of transformations, the previous two phases of the protocol, they begin this protocol with X^{n_s}, Y^{n_s} strings, the result of transformations in the previous two steps. For the input sequences X^{n_s}, Y^{n_s} , it is

$$P\left(X^{n_s} \equiv Y^{n_s}\right) = 1$$

That is, these arrays are identical to the probability of one. Using this fact and the publically known function, Alice and Bob form a sequence S_{AB} that they will use to perform a cryptographic key according to a pre-arranged and publically known procedure. Also, Eve using her starting sequence Z^n and the information gathered during Alice and Bob's communication through a publically available channel forms her sequence S_E which is an approximation of the sequence S_{AB}.

The previous steps of this process were conducted in such a way that constant care was taken to limit the amount of information flowing towards Eve, and therefore it can be shown that

$$I\left(S_{AB}, S_E\right) = 0.$$

which means Eve doesn't know anything about the string to perform the key between Alice and Bob. In other words, Eve knows for sure that she is, $S_{AB} \neq S_E$ but she have no information in what positions of discrepancy there is. Now Alice and Bob perform the cryptographic key k as

$$k = f\left(S_{AB}\right)$$

using, as mentioned earlier, a pre-known function f.

If chosen f correctly, her knowledge of Eve means nothing because of $S_{AB} \neq S_E$ and inability to predict the propagation of errors during computation because her error positions are unknown, $f\left(S_E\right) \neq k$ Detailed formalization and evidence of the described properties are based on the Renyi entropy and its derivatives, collision and minimal entropy. A fully formal treatment, proofs and detailed references can be found in [32].

Chapter 5

EEG based cryptographic keys agreement protocol

As a result of recent developments in computation processing technology, quantum computing, the information theory based advance towards information security has garnered a fresh round of attention. A cryptographic system is able to offer absolute security, ITSM model, of communications if and only if the uncertainty (entropy) of protected message is not greater than the entropy of the secret key, [7]. This is the core result of this technique, and it is easy to formulate. It has been established that systems that are designed in this manner are unaffected by the infinite computational power that potential adversaries possess, and consequently, to cryptanalysis that is based on that computing power, [34].

It is therefore evident that it can be argued that we have reached an age where detecting uncertainty of every imaginable nature, origin and collection point, with maximum entropy in a statistical sense, has become a priority for cryptographic key generation processes.

In this particular setting, the cornerstone results contained in the papers of Ahlsvede and Csiszar [12], Maurer [11], and Csiszar and Narayan [35] are particularly noteworthy and worthy of special attention. The fundamental concept behind this method is to extract signals that are mutually correlated and have entropies that are sufficiently high.

On the basis of the uncertainty source location, the two techniques that are described below can be differentiated from one another [11]:

(i) The process of extracting randomness from sources that are not depen-

dent on communication channels is referred to as the source model.

(ii) The process of extracting randomness from the transmission channel is referred as the channel model.

In this research, we investigate the feasibility of obtaining cryptographic keys from electroencephalography (EEG) signals by employing a technique that is based on the source model. In this particular instance, the electroencephalogram (EEG) signals are captured by means of the 14-channel EMOTIV EPOC+ wireless EEG headset [36], [37]. It is for two reasons why the electroencephalogram (EEG) was selected as the source of unpredictability.

To begin, the Secret-Key Agreement (SKA) is responsible for reliable and secure encryption keys agreement procedure in the symmetric encryption systems for participants who do not hold in advance disseminated secret symmetric cryptographic keys. This is the primary function of the SKA. In certain military applications, when keys cannot be delivered by direct physical distribution, or in situations involving undercover and special operations, where players do not have in advance generated and delivered cryptographic keys, this is a pattern that is commonly observed. The putting apart of functional blocks constitutes a foundational aspect in the design of professional information security systems. This is due to the fact that it reduces the likelihood of compromising the entire system by compromising a single component of the system. As a result, a SKA framework has to be autonomous of cryptographic and broadcast communications modules, which suggests that the SKA channel demonstrate ought to not be utilized.

The utilization of humans' biometric characteristics would be of significant benefit because it would eliminate the need for an extra arbitrary source as well as the dangers and costs related to it (research, generation, quality control, secure capacity, etc.).The SKA source model show is the as it were one that's still accessible

Second, it is crucial to examine the commercial availability, robustness, and usefulness of the sensor system that corresponds to the signal when selecting the biometric feature. In the set of possible biometric signals, which incloude gait, motion, electromyography (EMG), electrocardiogram (ECG), and electroencephalogram (EEG) (see review [38]), the EEG is distinctive because to its large entropy content, as well as the commercial availability of EEG sensors that are of the needed quality and resilience. There is no doubt that the primary function that EEG sensors and processing systems play in

contemporary Human-Computer Interface systems is the primary factor that drives their availability (see review [39]). Regarding this particular aspect, the EMOTIV EPOC+ system satisfies all of our necessities.

In the second section, we present the arguments that support the notion that a group of individuals who have been subjected to a particular mental task can be seen as an example of a single discrete memoryless source (DMS). During the course of our research, we captured the electroencephalograms (EEG) of 76 participants in an asynchronous manner as they were working through the Wisconsin card sorting test (WCST), Figure 5.4, [40], [41]. It is possible to substitute the WCST test with any other mental job, such as reading the specified text or viewing the selected image [42]. The WCST test has been chosen arbitrarily.

We conduct an analysis of the probabilistic and information theoretical properties of the collected information. Additionally, we determine the characteristics that are the most significant for every single round of the suggested SKA. These phases are referred to as advantage distillation (AD), information reconciliation (IR), and privacy amplification (PA).

An extended experiment was conducted in which secret keys were obtained for every possible communication pair among 76 participants ($76 * 75/2 = 2 = 2850keys$). The experiment was conducted for three different sorts of eavesdroppers, which are referred to as Eve: "Super evil Eve", "Medium evil Eve", and "Uninformed Eve".

Eve has access to a wide variety of information in the past, and these sorts cover the full spectrum. Following this, we will proceed to compare the results that were achieved regarding the characteristics of analogous systems that have been published in the currently available academic literature. The features of the suggested SKA are discussed along with the implementation possibilities that could be used in practice. Finally, in the seventh section, we examine the many methods that could be utilized to boost the rate of the secret key.

The conclusion incloudes a discussion of several unresolved concerns as well as the identification of a group of algorithms that are derived from the, commonly referred to, Data Exchange problem [43], [44]]. These algorithms are used to generate and distribute secret keys. Putting this strategy into practice in conjunction with the SKA system that was discussed in this article will be the focus of our further study.

5.1 Virtual communication channel on the Wisconsin Card Sorting Test

5.1.1 The Strategy of Sequential Key Distillation

A physical signal collecting samples model for SKA is depicted in Figure 4.4, which depicts a scenario where three individuals, namely Alice, Bob, and Eve, view output of Common Randomness Source (CRS) over DMS environment. Every one of them is given their own individual collection of observations. Each of Alice, Bob, and Eve's observations will be denoted by the letters X, Y, and Z, respectively. Although it is possible that all three parties are aware of the statistics regarding DMS, it is presumed that the three parties do not have influence over it. The objective of Alice and Bob is to reach a consensus on a secret key K, which will be based on their outputs from Common source of Randomness, X and Y, and ensure that Eve doesn't know anything about it beside her knowledge of Z. Within the context of the SKA scenario, a publically available transmission channel which allows access to all parties, inclouding Eve, so that Alice and Bob can mutually exchange messages with one another and share information. There is a widespread presumption that this public available channel is authenticated, which indicates that there is no possibility of impersonation.

A four-stage Sequential Key-Distillation (SKD) protocol is defined by the rules, which stipulate Alice and Bob to perform defined calculations over messages exchanged via a publically available channel and then reach a consensus on the secret key for the approach [11]:

i. *Randomness sharing.* Output from CRS is observed by Alice, Bob, and Eve that there are n bit usage $DMS(XYZ, P_{XYZ})$, where P_{XYZ} is the joint probability of the random variables X, Y and Z being considered.

ii. *Advantage distillation.* In the event that it is required, Alice and Bob will communicate with one another through a publically available channel in order to process their CRS outputs and "distill" the aspects of their observations wherein they are superior to Eve.

iii *Information reconciliation.* Alice and Bob communicate with one another on the publically available channel to interpret their findings and reach a consensus on a binary sequence that they both agree upon.

iv. *Privacy amplification.* There is a public agreement between Alice and Bob over a deterministic function that they would use to generate the secret key by applying it to their shared sequence.

As an effectiveness measure we use secrecy capacity of a publically available channel defined as the maximal rate at which information can be accurately conveyed among authorized parties. This capacity guarantees an arbitrarily low rate of reception of this information by an eavesdropper. Similary, term "secret key capacity" refers to the longest string length that can be exchanged in the eavesdropper presence. Formally, definition is as follows

$$C_k = min\left\{I(X;Y), I(X;Y \mid Z)\right\} \tag{5.1}$$

with $I(X;Y)$ represent the volume of information that exists among X and Y, and $I(X;Y \mid Z)$ represents the information volume existing that exist among X and Y that is under condition Z. In the case that Eve is independent of Alice and Bob, which means that Z is statistically independent of X and Y, the capacity of the secret key is equal to.

$$C_{k_{\max}} = I(X;Y) \tag{5.2}$$

This is a specific case $I(X;Y)$ makes up C_k maximal.

One of the benefits of the SKD technique is that it has been demonstrated to be successful in achieving all secret key rates that are lower than the amount of secrecy capacity C_k, in addition to its explicit practical implementation [11].

We propose the application of the SKD strategy to generate random sequences from $DMS(XY, P_{XY})$, where observations X and Y represent six-dimensional performance metrics signals obtained from the EMOTIV EPOC+ EEG headsets, which were worn by two subjects asynchronously engaged in the same mental task (see Figure 5.1 for more information). This is based on the strong theoretical result that we obtained. In the context of random variables X and Y, the term "P_{XY}" refers to the joint probability measure.

When Figure 4.4 and Figure 5.1 are compared, it is possible to observe that the CSR in Figure 4.4 has been replaced by the challenge of finding a solution to the WCST. In contrast to the traditional scenario depicted in

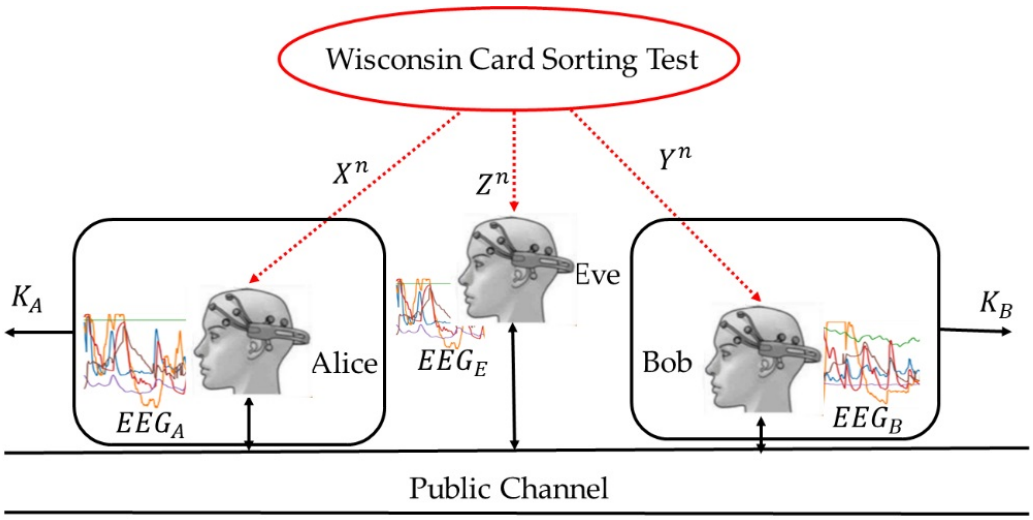

Figure 5.1: Secret-key establishment model using EEG

Figure 4.4 , in which the correlation of the observed data is brought about by the features immanent to physical phenomena, the resulting stochastic strings (X, Y, Z) depicted in Figure 5.1 are correlated because the individuals who participated in the test had mental processes that are comparable to one another. This structure of correlation is invariant to the following:

– time and location of the examination, as well as

– the individuals who were put through the test,

This makes it possible to acquire EEG signals in an asynchronous manner. In practical scenarios where synchronization is difficult to establish or would need more SKD system complexity and/or resources, this trait is of additional significance. It is of special value in certain situations.

5.1.2 Laboratory setting for capturing the electroencephalograms of the individuals who took the WCST test

For the purpose of this study, the data were gathered during sessions in which the participants were utilizing a variety of computer programs for neuropsychological testing, one of which being the Wisconsin Card Sorting

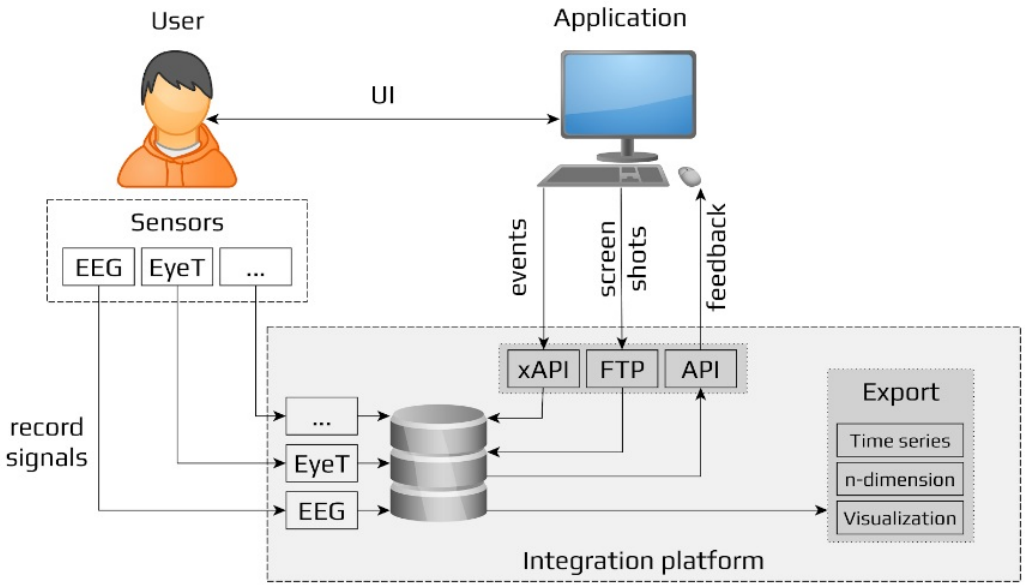

Figure 5.2: HCI monitoring and analytics platform (HCI.MAP), [45].

Test. Electrical brain waves recording and eye-tracking devices are among the sensors that are utilized throughout the sessions. In addition, the motions of the mouse and the strokes on the keyboard were recorded. The collection and synchronization of data (signals, application events, screenshots, and so on) was accomplished with the assistance of the Human-Computer Interaction Monitoring and Analytics Platform (HCI-MAP) [45], the architecture of which is depicted in Figure

The EMOTIV EPOC+ device, which is a wireless EEG headset with 14 channels and was created for measuring the cortical activity of the brain [46], was used to gather the electroencephalography signals. The application makes use of A/D conversion in conjunction with sequential sampling at a sample rate of 128 Hz. For the purpose of removing interference from the electrical power supply, its output frequency range is flat from 0.2 to 45 Hz and has digital notches at 50 Hz and 60 Hz. By utilizing a typical WiFi connection operating at 2.4GHz, the device was successfully linked to the HCI.MAP platform. In order to extract performance data for six different cognitive states—stress, engagement, interest, excitement, focus, and relaxation—the software that is offered by the manufacturer makes use of algorithms that are already integrated into the software.

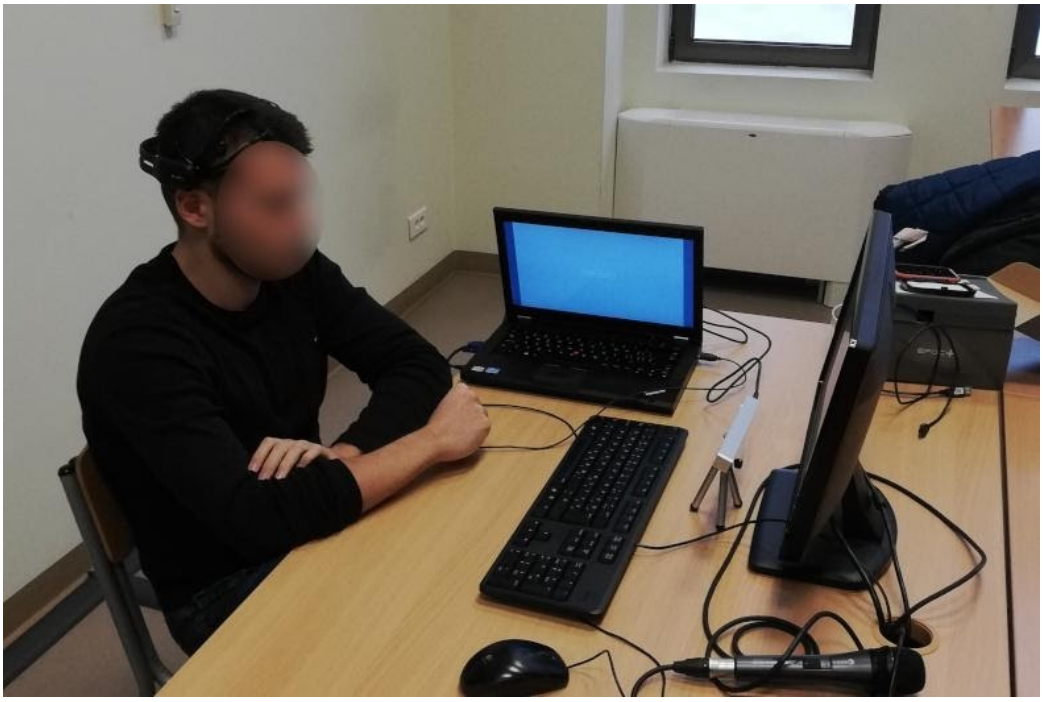

Figure 5.3: An Experimental environment with EEG and eye-tracking sensors enabled, [45].

For the purpose of assessing cognitive abilities that are referred to as executive functions, the Wisconsin Card Sorting Test (WCST) is utilized. Fundamental executive functions incloude the ability to focus attention and flexibility, as well as self-control and working memory. The examination can also be used to evaluate more sophisticated mental abilities, such as planning, reasoning, and problem-solving, all of which require the simultaneous use of a number of fundamental executive functions.

The each session event was videotaped as part of a study that inclouded 76 participants, all of whom were between the ages of 15 and 25 and were chosen using a random sampling procedure. The participants were aware of the protocol that would be followed for the research, which inclouded the use of the sensors, and they willingly agreed to take part in the test. Additionally, they were aware that the examination would be carried out in a manner that would ensure their anonymity; their data were merged together and omitted all identifying information. Gender, age, and educational level were the only

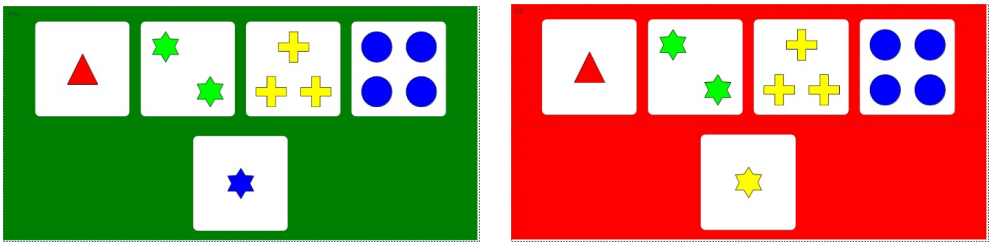

Figure 5.4: WCST application screen after selecting correct (left) and incorrect (right) cards.

pieces of personal information that were saved. All of the principles outlined in the Declaration of Helsinki were adhered to by the institutional ethics committee, which gave its approval to this research. It was a computerized version of the exams that were administered to the participants. Additional sensors were provided via the computer's mouse and keyboard to the computer. In order to be eligible for participation in the study, participants had to meet the medical requirement of not having any neurological or mental illnesses, inclouding addiction.

5.1.3 Acquisition of electroencephalogram (EEG) signals

Real-time measurement and recording, resulted in six dimensional time series sample obtained for each test partaker, where each measured feature express a different performance characteristic, more specifically denoted by Interest (i.e., emotional valence, attractiveness, or averseness of the task at hand), Engagement (or boredom, in negative valence, reflecting the mental workload), Excitement (arousal, emotional intensity), Stress (frustration), Relaxation (meditation) and Focus (attention) [47]. Due to the fact that EMOTIV EPOC+ is a proprietary software, the precise method that it employs in order to compute these performance indicators has not been revealed in its entirety. It was hypothesized that Emotiv Inc. established this system based on robust experimental investigations that had involved volunteers for each of the cognitive and emotional states that were described above. Prior to the measurements, the respondents were instructed by seasoned psychologists to reach varying degrees of the mental state that was being targeted.

Additionally, during the course of the studies, additional physiological measurements were gathered. These measurements inclouded eye tracking, pulse, respiration, blood pressure, blood flow, skin impedance, and blood pressure. Based on the findings of numerous independent researches [37], [47], it has been demonstrated that the EMOTIV EPOC+ EEG system is an effective method for assessing emotional states.

Because of the cross-correlation structure of these six metrics, it is very important to highlight that they do not have an effect on the performance of the suggested SKA. This is not because they are called after the emotional states that were described, but rather because of their origin. When it comes to the 14th channel EEG of each participant, it is important to note that these performance measurements comprise of six fixed changes that are executed in a consistent manner.

Alice, Bob, and Eve were chosen at random from among all 76 people who participated in the test, and their recorded signals are displayed in Figure 5.5.

After the signal acquisition, the dimensionality reduction is performed, which reduces the number of dimensions from six to one. This process ultimately results in a univariate time series for each partaker. Therefore, the transformation that is being applied is a straightforward serialization. This is because it is essential to maintain the dependency structure that exists among the partakers. To be more specific, a buffer is responsible for accepting a six-dimensional measurement vector at each sample point, and then sending out its components in a sequential manner. With regard to Alice, Bob, and Eve, the one-dimensional signals that were produced as a result of Figure 5.5 are displayed in Figure 5.6.

From Figure 5.6, we can concloude the pre-processing modification that was described earlier has, in fact, been successful in preserving the intrinsic correlation structure of the collected signals of the subjects. From this point forward, the signal that has been preprocessed will be named "principal EEG sample."

Quantization of the principal EEG sample is the next stage in the preprocessing process. There have been numerous investigations into this issue in the published works, however the majority of these investigations have focused on the sequential key distillation procedures for the channel model [15], [48], [49]. Both [50] and [51] demonstrated that there is a significant distinction between the discrete and continuous sources (see to Remark 5 in [51] for further information). For discrete sources, the upper limit of

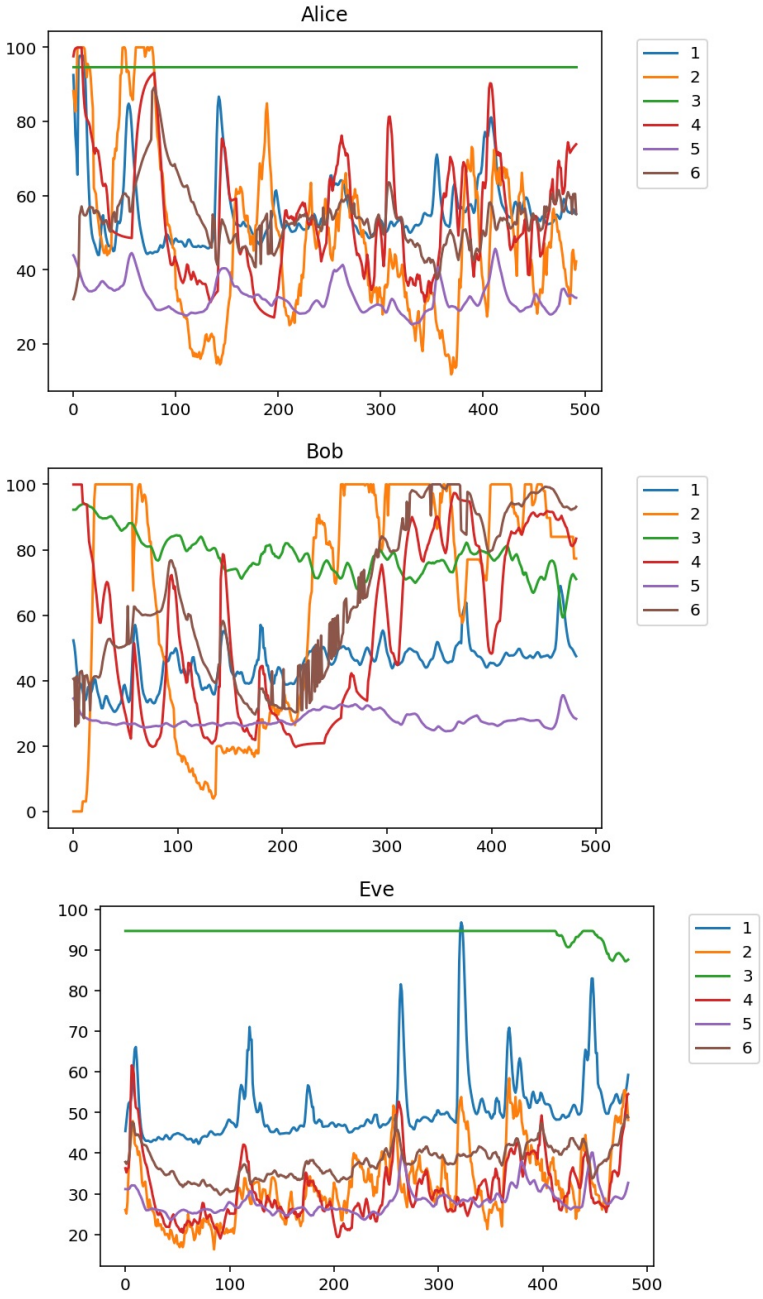

Figure 5.5: An illustration of performance metric signals randomly chosen from the entire pool of 76 test participants for Alice, Bob, and Eve.

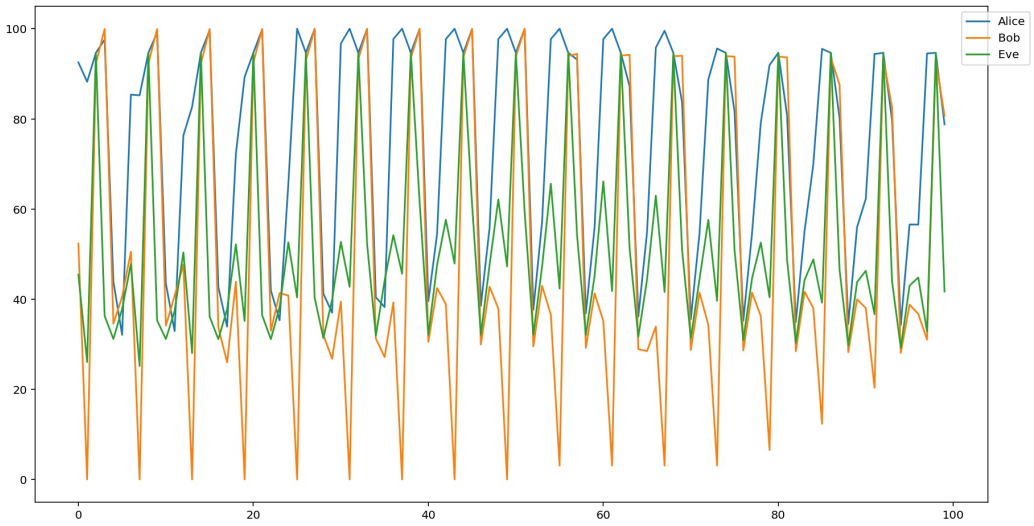

Figure 5.6: xxxxDimensional reduction of samples from Figure 5.5 from six to one

the secret key extraction rate can be achieved without quantization when the data rate across a publically available channel exceeds $H(X|Y)$. This is possible by employing the privacy amplification (PA) process and Slepian-Wolf coding, [52]. This is possible even if quantization is not used. On the other hand, as far as continuous Gaussian sources are concerned, the top limit cannot be reached for any finite data rate when it is transmitted via a publically available channel. According to the findings presented in [53] (Proposition 5.6), it was demonstrated that if X_q is a quantized version of X that is uniformly and finely sufficiently quantized, then the mutual information $I(X_q;Y)$ approaches the initial $I(X;Y)$ exponentially rapidly as the data rate on a publically available channel gets higher. Because of this, advanced quantization systems, such as the TCVQ (Trellis Coded Vector Quantization scheme), are only applicable in situations where there is minimal messages exchange over a publically available channel. Due to the fact that the primary objective of this study is to conduct an experimental confirmation of the suggested concept, we decided to go with the most straightforward scalar uniform quantization. This was done in order to circumvent the limitation of the publically available channel data rate.

According to the Shannon definition of the block entropy is as follows, [7]:

$$H_n = -\sum_{a_1,a_2,\ldots,a_n} P\left(a_1, a_2, \ldots, a_n\right) \log_2 P\left(a_1, a_2, \ldots, a_n\right) \tag{5.3}$$

and the probability that the vector $(a_1, a_2, \ldots, a_n)$ will appear as the output of an stochastic source is $P\left(a_1, a_2, \ldots, a_n\right)$. The entropy in question is referred to as the $n-$block entropy. Quantity that is referred to as the normalized block entropy, NBE, is numeric value

$$\frac{H_n}{n}$$

and its asymptotic value

$$\lim_{n\to\infty} \frac{H_n}{n}$$

is referred to as the Shannon/block entropy rate. The entropy of a finite sequence x whose length is N, in practice is a subject that piques our interest.. There is a possibility of making an estimate of $P\left(a_1, a_2, \ldots, a_n\right)$ based on the vector frequencies that are noticed in x if one considers a finite sequence x to be representative output from a observed source of information. In the case that x is binary string, possible binary patterns belongs to the set of all binary strings of the length n. Normalized block entropy is equivalent to volume of information bought by one bit of x.

The variation in the normalized block entropy of the principal EEG sample that was investigated is depicted in Figure 5.7. This variation is a function of the amount of bits per sample that were quantized using a uniform quantizer. The values of the block length change ranged from 1 to 20, and this function was calculated for those different values.

As can be seen in Figure 5.7 an increase in the number of bits per sample (also known as word length or bit depth) results in an initial increase in normalized block entropy, which is followed by a fall in that entropy. There is a correlation between an increase in the block entropy in the range $n_b = [1, 7]$, where n_b is the number of bits per sample, and a more precise depiction of the informational substance of the principal EEG sample. According to one interpretation, the subsequent decline in normalized block entropy in the range $n_b = [8, 16]$ might be understood as over-quantization, which results in the introduction of more redundancy in the principal EEG sample. It has been observed by a number of authors, the most notable of whom is [54],

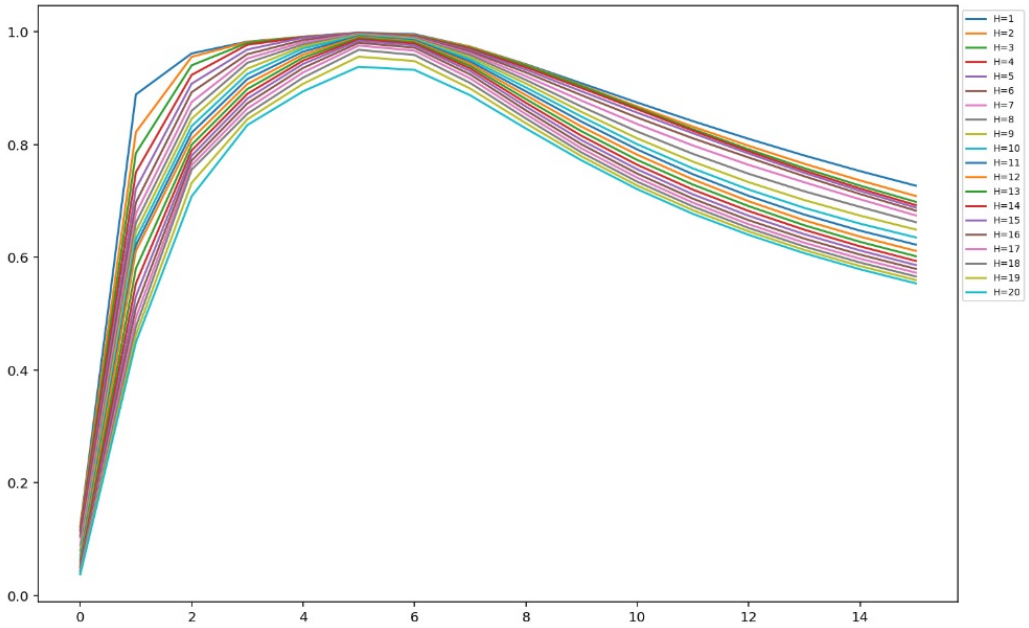

Figure 5.7: The normalized block entropies of the principal EEG sample were determined via uniform quantization, and their values were plotted regarding a number of sampled bits.as a function of the number of bits per sample. Each curve represents a certain block length value ranging from 1 to 20.

that the pace of secret key extraction may be increased by over-quantization. Taking into consideration this phenomenon, we made the decision to create a system that would function in two incarnations, one whose quantization value is: $n_b = 5$, which assumes the under-quantization model, and one whose quantization value is $n_b = 10$, which assumes the over-quantization model. We then proceeded to study the influence that these two quantization values would have on the overall performance of the system.

5.2 System description for SKA based on the principal EEG origin

5.2.1 Characteristics of the original source regarding of statistics and information theory

The principal EEG origin for $n_b = 5$ and $n_b = 10$ are discussed in Figures 5.8 and 5.9, respectively, and their fundamental properties are displayed. Both the histogram of the signal sequence length for each of the 76 people who participated in the test and the histogram of the NHDs D_h for each of their pairs are inclouded in the fundamental characteristics. The expression that represents the NHD between two binary sequences of the same length, identified as X and Y, is as follows:

$$D_h = \frac{\text{number of non-match bits}}{\text{number of bits compared}} \tag{5.4}$$

For sequences of varying lengths, the value of D_h is determined by applying the following expression:

$$D_h\left((x_1, x_2, ..., x_n), (y_1, y_2, ..., y_m)\right) = D_h\left((x_1, x_2, ..., x_p), (y_1, y_2, ..., y_p)\right), \tag{5.5}$$
$$p = \min\{m, n\}$$

In this manner, we were able to reduce the amount of data that was rejected during the SKD operation and the evaluation of its performance on each and every pair of participants. Assuming that the principal EEG source sequences are made up of binary iid random variables, the conditional entropy $H(X|Y)$ and the mutual information $I(X, Y)$ are calculated as follows:

$$\begin{aligned} H(X|Y) &= h_b\left(D_h\left(X, Y\right)\right) \\ I\left(X, Y\right) &= H(X) - h_b\left(D_h\left(X, Y\right)\right) \end{aligned} \tag{5.6}$$

the binary entropy function is denoted by the symbol h_b and is equal

$$h_b\left(p\right) = -p \cdot \log_2 p - (1-p) \cdot \log_2\left(1-p\right) \tag{5.7}$$

In light of the fact that the function h_b is rising in a monotonic fashion inside the interval $[0, 1/2]$, the maximal extraction rate of secret keys C_k,

as described by (5.1), may be accurately measured by $D_h(X, Y)$ for a predetermined quantity of information that Eve possesses on sequences X and Y.

Recall that the histogram of the mutual NHD of random and entirely independent sequences is narrowly concentrated around the value of 0.5. This is the situation when the sequences are completely independent of one another. For the purpose of establishing the validity of this assertion, let us consider the binary random variable D_i that represents discrepancy of two binary strings,X and Y, of length p in position i. This set of p random variables is independent, meaning that they have an equal probability of either 0 or 1, i.e. $P(D_i = 0) = P(D_i = 1) = \frac{1}{2}$. As a result of the linearity of mathematical expectation

$$E(D_1 + D_2 + \cdots D_{p_i}) = E(D_1) + E(D_2) + \cdots + E(D_{p_i}) = \\ = \frac{1}{2} + \frac{1}{2} + \cdots + \frac{1}{2} = \frac{p}{2}$$

In light of this

$$E(D_h(X, Y)) = \frac{1}{p} \cdot E(D_1 + D_2 + \cdots D_{pi}) = \\ = \frac{1}{p} \cdot \frac{p}{2} = \frac{1}{2} \tag{5.8}$$

A movement toward smaller NHD, that is, smaller disparities among the samples, may be detected by the comparation of the right side histograms in Figures 5.8 and 5.9. This change can be seen by comparing the two figures. The fact that this is the case demonstrates once again that over quantization brings about additional dependecies to the collection of samples of the original source. Given that the EMOTIV EPOC+ device has a sampling rate of two samples per second, Consequently, the test duration can vary from eighty-three to five hundred seconds, with a typical value of approximately two hundred and fifty seconds.

5.2.2 A model of an eavesdropper

Due to the experimental assessment being conducted on a participant group, we can distinguish three common scenarios from an eavesdropper's standpoint (Eve), based on the level of prior information available about the primary source.

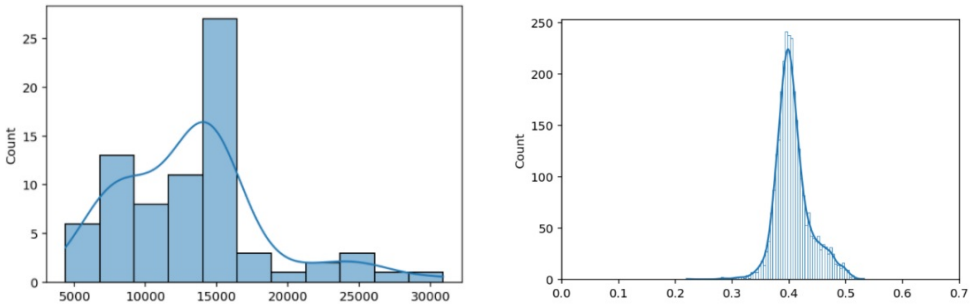

Figure 5.8: Histograms are displayed using uniform coding ($n_b = 5$) with 5 bits per sample.The distribution of EEG sequence lengths for each participant is shown by the left histogram, and the distribution of NHD across all pairs is shown by the right histogram. The sequence has 1,006,560 bits in volume. The NHD' mean and dispersion are given as $0.41 + / - 0.036$

A) Not only does the eavesdropper know who Alice and Bob are, in addition they have access to all of the EEG samples of the people who participated in the test, with the exception of the signals that belong to Alice and Bob. Moreover, the eavesdropper possesses knowledge about which signals provided by the participants closely resemble those submitted by Alice and Bob.. Therefore, the attacker has the ability to pick up Eve, who is the person whose sample is closest to Alice and Bob in terms of the NHD, for each pair of Alice and Bob by using adaptive selection. This establishes the most challenging conditions for extracting secret keys, where Eve should have no information whatsoever. This holds true both in theory and practical application. Because of this, the term "Super Evil Eve" (SE) is commonly used to refer to this particular type of eavesdropper.

B) The eavesdropper lacks knowledge about the identities of Alice and Bob, leading him to select Eve in a manner where her position is equidistant from all participants in NHD. This is comparable to the centroid of a cluster encompassing the whole population. Consequently, we informally refer to this Eve as the "Medium evil Eve" (ME). In the case of the examined primary source, ME correspond to subject No. 62, as illustrated in Figure 5.10.

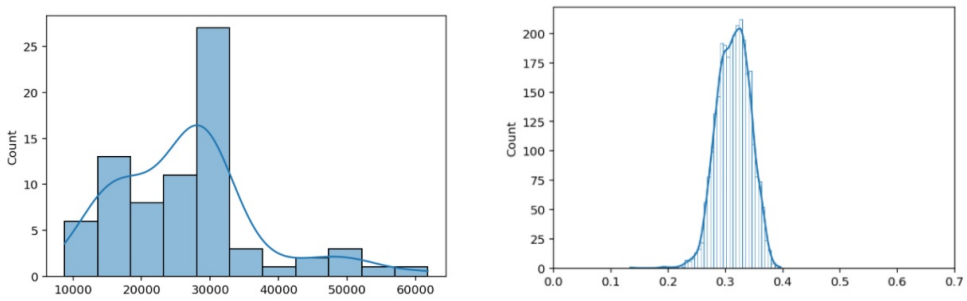

Figure 5.9: Histograms are displayed using uniform coding ($n_b = 10$) with 5 bits per sample.The distribution of EEG sequence lengths for each participant is shown by the left histogram, and the distribution of NHD across all pairs is shown by the right histogram. The sequence has 2 013 120 bits in volume. The NHD mean and dispersion are given as $0.41 + / - 0.036$

C) The eavesdropper possesses no specific details about the primary source, aside from its composition of EEG signals acquired through the EMOTIV EPOC+ device. In this scenario, the optimal tactic for the channel observer involves recording their EEG signal and participating in the protocol as Eva. We informally label this eavesdropper "Uninformed Eve" - UE. During the trials, UE refers to an external subject who is not part of the test group. Their EEG is recorded while they observe a single image, notably the recreation of the famous icon "White Angel" from the Serbian medieval monastery Mileševa. [42], [55] for 768 seconds. In the cluster analysis performed, this subject is denoted by numeral 76, as shown in Figure 5.10.

Figure 5.10 depicts the dendrogram that was generated using Ward's approach [56] for the purpose of conducting hierarchical cluster analysis on the primary source signal. An input matrix, which is created by the NHD, is what the clustering algorithm takes into consideration. Despite the fact that his EEG readings were the consequence of a completely different mental task, It can be noted that the individual labeled UE does not exhibit substantial differences from the other participants in the test.

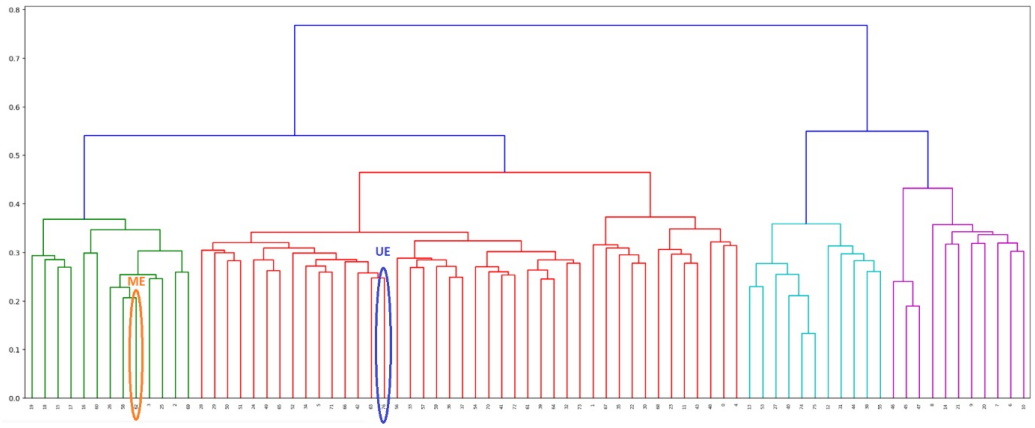

Figure 5.10: A dendrogram, created using Ward's approach, illustrates the hierarchical cluster analysis of the primary EEG data from all participants in the test. The entities "Medium Evil Eve" and "Uninformed Eve" are represented by their acronyms and distinguished by orange and blue encirclements, respectively.

5.2.3 Architecture of the SKD system that is being proposed

The fundamental structure of the SKD system that has been proposed is shown in Figure 5.11. The ultimate objective for the proposed system is to guarantee that the authorized participants in the procedure, Alice and Bob, at the end have final symmetric encryption keys that are identical to one another, denoted by the equation $K_A \equiv K_B$ with probability close to one. Furthermore, it is crucial that Eve's key K_E doesn't reveal any information about the keys of the participants. Eve is aware of all the components of the system as well as all of the parameters of the individual sub-blocks, which is in accordance with the fundamental Kerckhoffs principle, which states that security is not obscurity [57], [58]. It is demonstrated in [11] that the most effective tactic for Eve would be to carry out the imitation game regarding Alice and Bob while communicating through a publically available channel. The next step, which comes after serialization and consistent quantization, is the distillation of advantages, followed by information reconciliation and privacy amplification. The implementation of PA is accomplished by employing a particular family of universal hash functions. As the system operating cycle comes to a close, Alice and Bob have a secret key that is identical to

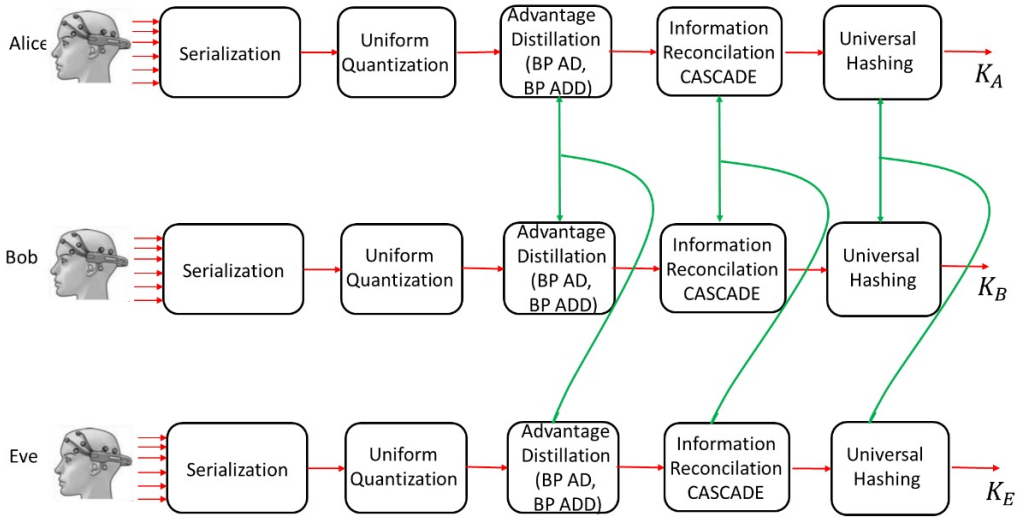

Figure 5.11: The proposed SKD system architecture is designed based on the asynchronous EEG data of the participants. Messages transmitted on the publically available channel are indicated by the color green.

one another, denoted by the equation $K_A = K_B$. On the other hand, Eve's key, denoted by K_E, lacks any pertinent information about them..

Advantage Distillation (AD)

In the typical scenario, it is essential to posit that Eve initially possesses an advantage over Alice and Bob. This means that the NHD between Eve's string and that of Alice (or Bob) is smaller than the NHD between Alice's and Bob's strings. The objective of the advantage distillation (AD) phase is for Alice and Bob to communicate messages over a publically available channel, thereby flipping the advantage in their favor.

Numerous advantage distillation (AD) algorithms have been documented in the literature, with the Bit pair (BP AD) protocol [59] and the more recent Bit pair advantage distillation/degeneration protocol (BP ADD) [17] being the most prominent. The main difference between these protocols is that, unlike BP AD, BP ADD not only reduces the NHD between Alice's and Bob's sequences but also increases the distance between Eve's and Alice's (Bob's) strings. [17].

The AD (Algorithm 1) and ADD (Algorithm 2) protocols are broken

down into their respective descriptions below. The i-th bit of sequences that were previously held by Alice and Bob, respectively, are referenced by the symbols Xi and Yi.

Algorithm 1 Bit Parity AD protocol:

1. Alice and Bob partition n_{AD_0} bits into pairs $(X_{2i+1}, X_{2i+2})$and (Y_{2i+1}, Y_{2i+2}) for $i = 0, 1, ..., \lfloor \frac{n_{ad0}}{2} \rfloor - 1$.

2. Alice and Bob calculate the values parity values of these blocks, $\left\{ X_{2i+1} \oplus X_{2i+2} \,|\, i = 0, 1, ..., \lfloor \frac{n_{ad0}}{2} \rfloor - 1 \right\}$ and $\left\{ Y_{2i+1} \oplus Y_{2i+2} \,|\, i = 0, 1, ..., \lfloor \frac{n_{ad0}}{2} \rfloor - 1 \right\}$

3. Alice transmits $\lfloor \frac{n_{AD_0}}{2} \rfloor$ calculated bits to Bob over the publically available channel. In the case that calculated values match Bob send acknowledge to Alice over publically available channel.

4. For the acknowledged pair i Alice keeps X_{2i+1} and Bob keeps Y_{2i+1} the next repetition.

Algorithm 2 Bit Parity ADD Protocol

1. For $k = 1, 2, ...$ Alice calculate $C_k = X_{2i-1} \oplus X_{2i}$ and transmits C_k to Bob;
   Bob calculates $D_k = Y_{2i-1} \oplus Y_{2i}$ and transmits it to Alice.

2. If $C_k \neq D_k$ Alice checks condition $X_{2i} = 1$ and if it holds Alice discards X_{2i-1} from X. In opposite case Alice discards X_{2i} from X. In the same way Bob discards $Y_{2i-1}Y_{2i}$ from Y.
   If $C_k = D_k$ Alice checks condition $X_{2i} = 1$ and if it holds Alice discards X_{2i-1} from X. In opposite case Alice discards X_{2i} from X. In the same way Bob checks condition $Y_{2i} = 1$ and if it holds Bob discards Y_{2i-1} from Y. In opposite case Bob discards Y_{2i} from his sequence Y.

The effectiveness of the BP AD and BP ADD protocols can be evaluated by examining Figures 5.12-5.15, illustrating the evolution of the distribution of corresponding NHD during the initial two iterations of these protocols. In Iteration 0 (depicted in blue), the initial distribution of NHDs for available primary source sequences is represented. A comparison of the mean values

of these distributions at the conclusion of the second iteration (depicted in green) for both the BP AD protocol (refer to Figure 5.12 and Figure **??**) and the BP ADD protocol (refer to Figure 5.13 and Figure 5.15) reveals a substantial advantage achieved by Alice and Bob with the BP ADD protocol over Eve. This observation will be further validated through a comprehensive experimental evaluation.

In practice, it is not uncommon for Alice and Bob to communicate via a line that is not authenticated, in which case they have to provide authentication mechanisms themselves. This is most commonly achieved by applying relatively short secret cryptographic keys, for example 128 bits long. Then, regardless of Eve's presumed initial advantage, Alice and Bob have an informational advantage already in the first iteration of any of the previous algorithms as a consequence of the following fact.

Let A, B and C be three random variables such that A, B are independent from C , then

$$H\left(A|B\right) \leq H\left(A|B \oplus C\right). \tag{5.9}$$

Nameli, if Alice and Bob use blocks whose length is equal to their secret key and each block xored with that secret key then it is easy to show that they have advantage against Eve.

Information reconciliation.

After the Advantage Distillation (AD) phase, Alice possesses significantly greater knowledge about Bob's sequence compared to Eve. To fulfill the objective of the Information Reconciliation (IR) phase, Alice needs complete and accurate knowledge of Bob's sequence. Every protocol within this category utilizes an iterative process involving communication in both directions via a publically available channel to identify and rectify any discrepancies that may arise in Alice's and Bob's sequences. Once faults are identified and corrected, Alice's and Bob's key vectors perfectly match, achieving the goal of this step.

Despite the availability of a range of IR protocols based on robust error-correcting codes, such as low-density parity-check codes [60], we opted for one of the most widely used and efficient IR algorithms—the Cascade protocol, initially proposed in [19]. Compared to more complex error-correcting techniques, this protocol is generally believed to provide Eve with substantially

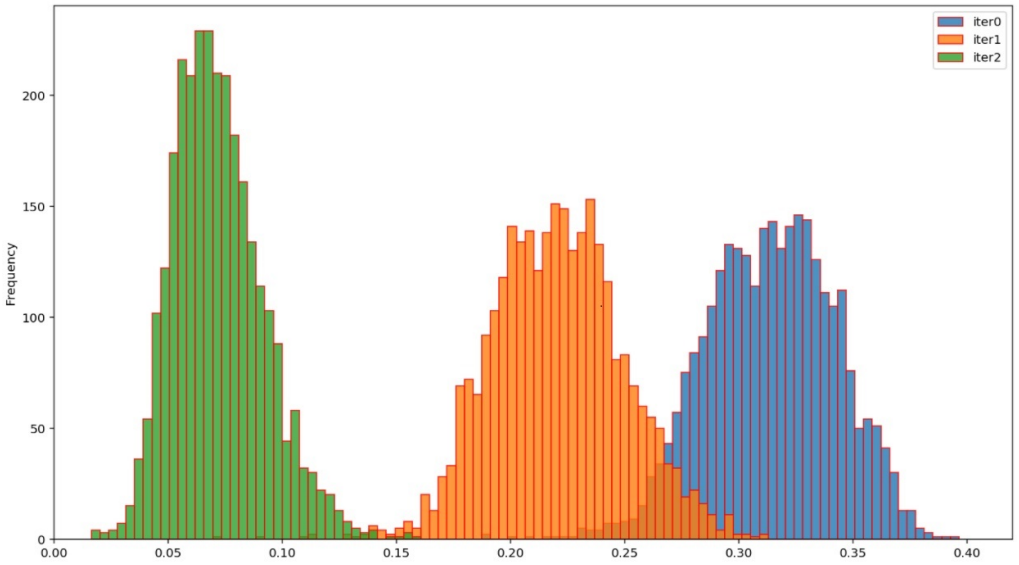

Figure 5.12: The changes in the distribution of NHD among the sequences of Alice and Eve during two iteratioA, B, s of the BP AD protocol with $n_b = 5$.

less information about the common sequence derived by Alice and Bob. Continuously refined and optimized, the Cascade protocol has found widespread application in the field of quantum key distribution. In this paper, we adopt an implementation detailed in [61] and the associated GitHub repository.

The Cascade information reconciliation technique requires several cycles to complete its process. During each round, the sequences of Alice and Bob are divided into blocks, and the parity of these blocks is compared, enabling the identification and correction of errors if they occur. Before execution, Alice and Bob agree on the number of iterations and the block size for the initial iteration.

Algorithm 3 Cascade protocol
INPUT: A,B %Alice and Bob sequences
OUTPUT: K %reconciled key

1. During the initial iteration, Alice and Bob partition their strings into blocks, and Alice then sends Bob the parities of all of her blocks

2. Bob computes his parities and uses Binary algorithm (Algorithm 4) for

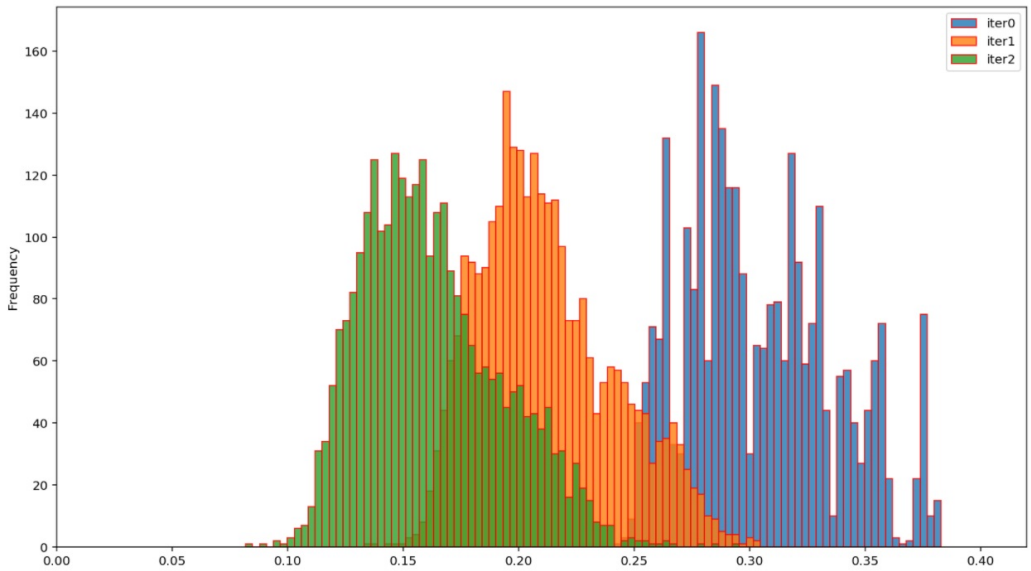

Figure 5.13: The changes in the distribution of NHD among the sequences of Alice and Eve during two iterations of the BP AD protocol with $n_b = 10$.

error detection and recovery.

3. In the beginning of the each and every other iteration, Bob is required to reshuffle the bits of his key and to do again steps 1 and 2. using enlarged blocks, actually doubled in size.

4. Because of bit value correction cascade effect on shuffled blocks in previous iterations arise and new session of Binary algorithm on previous blocks is applied.

5. Cycle trough steps 3. and 4. as long as the specified number of iterations is reached.

Algorithm 4 Binary algorithm
In the event those blocks of keys A and B exhibit contradictory parity:

1. Alice partition her block in two halves and delivers to Bob the parity of the first halve.

2. In the same manner, Bob divides his block, and then compares his first halve calculated value with Alice's in order to discover which part of the block has an uneven number of different positions.

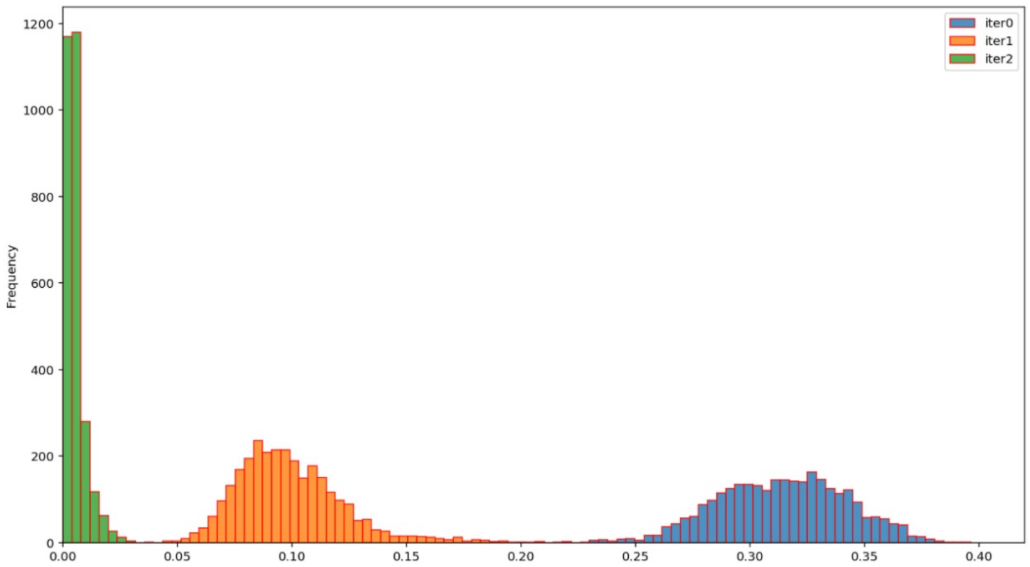

Figure 5.14: The changes in the pattern of distribution of the NHD among Alice's and Bob's sequences across two iterations of the BP ADD technique with $n_b = 5$.

3. Repeat previous steps as many times as necessary until a mistake is discovered.

Example 1, Example 2, and Example 3 each explain how the AD protocol, the ADD protocol, and the Cascade protocol operate, respectively.

The value of A is equal to 01

Privacy amplification (PA)

Throughout the implementation of any Information Reconciliation (IR) protocol, Eve can acquire partial knowledge about the shared sequence inffered by Alice and Bob by monitoring the publically available channel. Consequently, the final step in the Secret Key Distillation (SKD) technique involves the application of a suitable transformation. This transformation aims to minimize Eve's knowledge to a negligible extent. Consider a hypothetical scenario where Eve is provided with information about the calculated value for each individual block forming the ultimate shared sequence during the execution of the Cascade protocol. From a cryptanalysis perspective, this

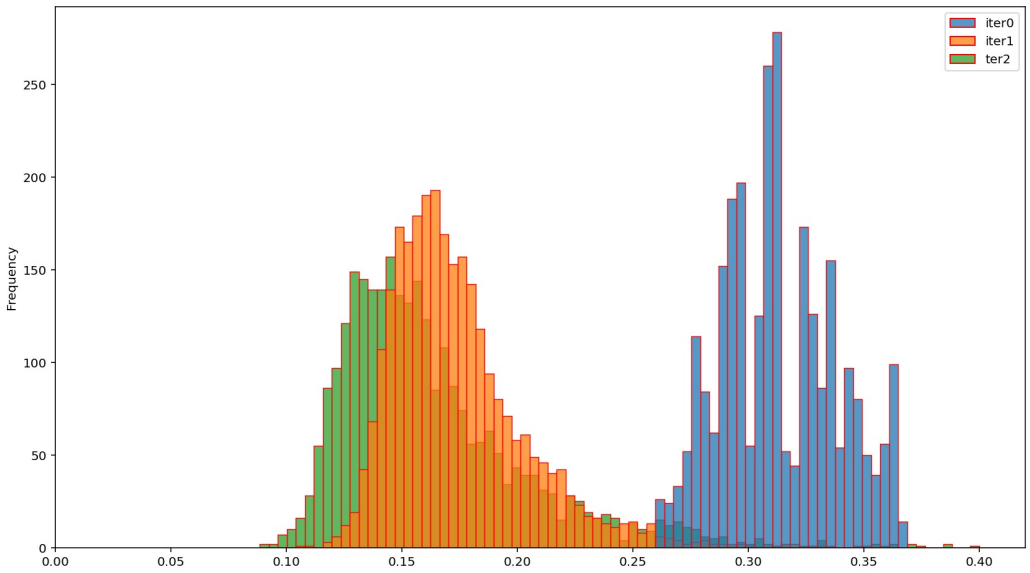

Figure 5.15: The progression of the distribution of NHD among the sequences of Alice and Eve throughout two iterations of the BP ADD protocol with $n_b = 10$.

situation resembles a form of algebraic attack, wherein the adversary constructs a system of linear equations by adding equation for each observed parity value over used unknown bit values of the Alice's and Bob's sequence.

A well-established technique known as the Leftover Hash Lemma [62] underpins the most common approach to constructing Privacy Amplification (PA) algorithms. This technique addresses the question of whether a cryptographic key of length n, of which the adversary is aware of the values of some t bits (where $t < n$), can still be used or if it should be discarded in favor of a new key. The answer is that a partially compromised key can be utilized, and through the application of the required transformation, a key of approximately $n - t$ bits in length can be generated, of which the adversary is largely unaware. It was demonstrated in [**?**] that the described transformation can be any hash function belonging to the universal class of hash functions (where k is the length of the output hash string).

The proposed SKD system underwent experimental evaluation, employing a universal class of hash functions, as expressed in the following.

Example 1 Bit Parity AD protocol

INPUT: A = 00101110001001101101110110000000010
B = 00111101110001010101111010011011 01

1: A = 00|10|11|10|00|10|01|10|11|01|11|01|10|00|00|00|10
B = 00|11|11|01|11|00|01|01|01|01|11|10|10|01|10|11|01

2,3: A = 00|~~10~~|11|10|00|~~10~~|01|10|~~11~~|01|11|01|10|~~00~~|~~00~~|00|10
B = 00|~~11~~|11|01|11|~~00~~|01|01|~~01~~|01|11|10|10|~~01~~|~~10~~|11|01

4: A = 011001010101
B = 010100011110

Example 2 Bit Parity ADD protocol

INPUT: A = 10111111001010110100101100110100 11
B = 01011100110010001001111111111111100

1: A = 10|11|11|11|00|10|10|11|01|00|10|11|00|11|01|00|11
B = 01|01|11|00|11|00|10|00|10|01|11|11|11|11|11|11|00

2: A = 10|~~11~~|11|11|00|~~10~~|10|11|01|~~00~~|~~10~~|11|00|11|~~01~~|00|11
B = 01|~~01~~|11|00|11|~~00~~|10|00|10|~~01~~|~~11~~|11|11|11|~~11~~|11|00

3: A = 011001010101
B = 010100011110

$$H = \left\{ h_M \quad : \; M \in GF\left(2\right)^{k \times n} \right\} \tag{5.10}$$

$$h_M = Mx \tag{5.11}$$

In this context, the symbol M represents a binary matrix with dimensions $k \times n$. All operations take place within a Galois field with 2 elements, denoted as $GF(2)$. During the execution of the Cascade protocol, the n_p denotes number of parity calculations requested over the publically available channel, then the relation $n_p > t$ is true. Here, t represent the number of bits of the common sequence between Alice and Bob that Eve be conscious of after the Information Reconciliation (IR) phase has conclouded. There is a potential for Eve to gain knowledge of one bit of the common sequence between Alice and Bob for each new parity inquiry. This is the worst-case situation. Considering that t equals n_p in this scenario, the dimension k of matrix M in equations (5.10) and (5.11) is as follows:

$$k = n - n_P \tag{5.12}$$

Example 3 Cascade protocol

Output of the previous iteration, S-1:

A = 0 1 1 0 0 1 0 1 0 1 0 1

B = 0 1 0 1 0 0 0 1 1 1 1 0

We need to detect and change unmatching positions, the yellow-colored zeroes and ones.

Permutation Iteration s
2 → 0
5 → 1
7 → 2
1 → 3
4 → 4
11 → 5
3 → 6
6 → 7
0 → 8
10 → 9
9 → 10
8 → 11

Next iteration, S:

B_S = 0 0 1 1 0 0 1 0 0 1 1 1 permuted sequence

B_S = 0 0 1 1 0 0 1 0 0 1 1 1

Request parity values for new blocks:

- 0 0 1 1 → unpermuted key positions = 2,5,7,0 → accurate parity = 1
- 0 0 1 0 → unpermuted key positions = 4,11,3,6 → accurate parity = 0
- 0 1 1 1 → unpermuted key positions = 1,10,9,8 → accurate parity = 0

Upper blocks of B in these round parities equal to the parities of comparing blocks, so in this round we did not correct any sequence element.

Shuffle Iteration N+0
5 → 1
9 → 0
1 → 2
01 → 3
7 → 4
2 → 5
8 → 6
6 → 7
0 → 8
3 → 9
00 → 01
4 → 00

Iteration S+1

B_{S+1} = 0 1 0 1 1 0 1 0 1 1 0 0 permuted key

B_S = 0 1 0 1 1 0 1 0 1 1 0 0

Request parities:

- 0 1 0 1 1 0 → unpermuted key positions = 5,9,1,01,7,2 → accurate parity = 1
- 1 0 1 1 0 0 → unpermuted key positions = 8,6,0,3,00,4 → accurate parity = 1

Both upper blocks of B in this round have different parity from accurate parity. One bit will be changed from each block.

B_{N+1} = 1 1 0 1 1 0 0 0 1 1 0 0 red-colored zeroes and ones are changed.

Performed changes in this round of the algorithm immediately affect blocks from previous round.

B_N = 0 1 1 1 0 0 1 0 0 1 1 0

Two upper blocks engaged in round S do not have accurate parity. So, we are able to change another two bits.

B_N = 1 1 1 1 0 0 1 0 0 0 1 0

Process will continue until all errors are corrected or the maximum number of rounds is reached.

The key length in the beginning and the number of exchanged messages regarding parity are both known (numbers n and n_P), which imply that that k is also simply calculated and therefore known. As a result, the hash functions that are given by (5.10) and (5.11)) may be computed and applied, which will result in the final shared encryption secret key becoming available. The Leftover Hash Lemma states that as a consequence of this, Eve's key K_E carries a minimal amount of information regarding the shared secret encryption key $K_A \equiv K_B$ that has been formed among the legitimate procedure participants.

The experiment outcomes

To assess the proposed Secret Key Distillation (SKD) system, two distinct principal EEG samples were employed. Two distinct quantization values were utilized to acquire results: $n_b = 10$ bits and $n_b = 10$ bits, respectively, for each sample. We explored two variations of SKD within the framework of advantage distillation. The first variant underwent testing with the Bit Pair Advantage Distillation (BP AD) algorithm, while the second variant was tested with the Bit Pair Advantage Distillation/Degeneration (BP ADD) algorithm. Subsequently, these algorithms will be denoted as AD and ADD.

For the evaluation involving the two quantization versions and all three types of Eve (EE, ME, and UE), tests were conducted on all $76 \cdot 75/2 = 2850$ unique pairings of individuals. The decision was made to set $n_a = 2$ as the number of iterations for the AD algorithm. It has been empirically demonstrated that this number of iterations was sufficient to achieve a substantial advantage for Alice and Bob over all Eve personalities. This choice represents a trade-off between increasing the advantage over Eve and minimizing the associated loss of sequence length at the output of the AD stage. The selected value for the parameter n_a reflects the outcome of finding this compromise.

In each quantization and advantage distillation scenario, the Cascade Information Reconciliation (IR) technique was employed, with the maximal number of iterations set to $n_c = 4$, and the starting block size for parity calculation set to $n_{block} = 8$. The operation of the cascade algorithm concloudes when the sequences of Alice and Bob are identical. The average number of iterations required to achieve this equality is represented by the value $\overline{n_c}$.

Here are some of the indications that were used to evaluate the success of the system:

- Final length of the key,

- The whole length of the keys that are final,
- Known as the key rate (KR),
- (IR) effectiveness,
- Finally, the NHD between Alice's and Eve's keys has been determined.
- The key agreement rate will be KA.
- Leakage rate (LR), as well as
- Mean entropy of the block.

The key rate is determined by the following:

$$KR = \frac{\text{total length of established keys}}{\text{total length of input sequence}} \cdot 100\,[\%] \tag{5.13}$$

In addition, the effectiveness of the information reconciliation is established as

$$IR_{efficiency} = \frac{m}{H\left(A \,|\, B\right)} = \frac{m}{n \cdot h_b\left(D_h\left(A, B\right)\right)} \tag{5.14}$$

In this equation, m denotes the total number of bits transmitted over the publically available channel during the Information Reconciliation (IR) phase, n represents the length of strings at the initiation of the IR phase, and h_b signifies the binary entropy function as defined in equation (5.7).

This relationship is defined by the connection among the communicated bits and the theoretical minimum, which was determined in the publication by Slepian in 1973, [63]. The ratio has a minimal value of 1, indicating an optimal IR procedure based on Slepian Wolf's optimal source coding of associated sources. Simultaneously, this value serves as a metric for the communication complexity of the IR protocol.

Next is the formula that gives the final normalized Hamming distance, $FNHD$:

$$FNHD\ \ (A, E) = D_h\left(K_A, K_B\right) \tag{5.15}$$

which represents the NHD between Eve's final key and the key shared by Alice and Bob. Ideally, these keys should be statistically uncorrelated.

If this is the case, then the predicted value of (5.15) is equal to 0.5. Using (5.8) from (5.15) applying definition of expectation value computing we obtain that it is equal to 0.5.

The key agreement rate, often denoted as KA, is determined by considering the expression.

$$KA = \frac{\text{number of succesfull key establishment } (K_A = K_B)}{\text{total number of attempts}} \cdot 100\,[\%]$$

The first table. Results of the AD protocol testing

Parameter	**na=2 nc=4 nblock=10 nb=10**			**na=2 nc=4 nblock=10 nb=5**		
Type of Eve	**EE**	**AE**	**UE**	**EE**	**AE**	**UE**
nc mean	2.27	2.26	2.27	2.57	2.59	2.56
Final key length (mean,std)	1301.55 ± 502.16	1290.53 ± 496.85	1301.76 ± 502.44	243.04 ± 138.77	242.48 ± 139.26	243.30 ± 137.28
Total length of final keys	3709416	3581223	3710007	587184	569341	587576
Key rate (KR) [%]	4.79	4.75	4.79	1.79	1.77	1.79
Leakage rate	0.0006 ± 0.0010	0.0006 ± 0.0010	0.0006 ± 0.0010	0.0058 ± 0.0198	0.0053 ±0.0170	0.0057 ±0.0168
IR efficiency	1.17 ± 0.05	1.17 ± 0.05	1.17 ± 0.05	1.17 ± 0.05	1.17 ± 0.05	1.17 ± 0.05
Final normalized Hamming (A,E)	0.4997 ± 0.0147	0.5005 ± 0.0149	0.4999 ± 0.0147	0.4997 ± 0.0527	0.5003 ± 0.0495	0.4988 ± 0.0487
Key agreement rate (KA) [%]	100	100	100	84.77	84.61	84.74
Mean block entropy (k = [1,20])	0.9989	0.9988	0.9989	0.9926	0.9924	0.9927

Table 1. Results for AD protocol

The second table. Results of the ADD protocol testing

Parameter	**na=2 nc=4 block=10 nb=10**			**na=2 nc=4 block=10 nb=5**		
Type of Eve	**EE**	**AE**	**UE**	**EE**	**AE**	**UE**
nc mean	1.37	1.37	1.37	1.92	1.93	1.92
Final key length (mean,std)	2454.28 ± 819.68	2435.94 ± 811.47	2454.09 ± 819.52	739.12 ± 297.97	743.11 ± 300.32	738.29 ± 298.51
Total length of final keys	6994706	6759745	6994143	2103530	2059898	2104116
Key rate (KR) [%]	9.04	8.96	9.04	5.44	5.44	5.44
Leakage rate (LR)	0.0003 ± 0.0005	0.0003 ± 0.0005	0.0003 ± 0.0006	0.0013 ± 0.0022	0.0012 ± 0.0024	0.0013 ± 0.0027
IR efficiency	3.63 ± 2.11	3.60 ± 2.11	3.63 ± 2.11	1.86 ± 0.63	1.85 ± 0.62	1.86 ± 0.63
Final normalized Hamming (A,E)	0.4998 ± 0.0106	0.4999 ± 0.0107	0.5005 ± 0.0109	0.5002 ± 0.0209	0.5000 ± 0.0200	0.5000 ± 0.0210
Key agreement rate (KA) [%]	100	100	100	99.86	99.89	100
Mean block entropy (k = [1,20])	0.9994	0.9994	0.9994	0.9979	0.9979	0.9979

Volume of information leakage to Eve per bit contained in K_A, K_B,common keys:

..

$$LR = I(X;Z) = 1 - h_b(D_h(A,E)) \tag{5.16}$$

The mean block entropy is defined by

$$\text{Mean block entropy } = \frac{1}{20}\sum_{k=1}^{20} H_K \tag{5.17}$$

with H_k denoting block entropy of block whose length is $k.$, defined in (5.3). This value describes the level of uncertainty on agreed keys. Figure 5.16 illustrates the variation in H_K for level K within the range from 1 to 20 for all six tested variations of SKD systems.

The level of indistinguishability from true random samples is usually checked by the proprietary designed statistical tests.

.

Table 3. Randomness test results of the AD and ADD key sequences based on the Statistical Test Suite developed by NIST.

	F	**BF**	**R**	**LR**	**FFT**	**S**	**AE**	**CSf**	**CSr**
AD	0.9114	0.5341	0.3504	0.5341	0.9914	0.7399	0.5341	0.7399	0.0668
ADD	0.0351	0.7399	0.3504	0.0088	0.7399	0.7399	0.1223	0.7399	0.2133

Table 3 displays the outcomes of randomness tests conducted on key sequences generated by the AD and ADD protocols. The randomness assessments utilize the Statistical Test Suite developed by the US National Institute of Standards and Technology NIST, as documented in [64]. Each experiment's outcome is expressed by the $P - value$, as indicated in Table 3. A specific test is deemed successful if the attained $P - value$ exceeds the threshold of 0.01. Based on the obtained results, it is evident that the key sequences produced by the AD and ADD protocols satisfy the established randomness criteria across all conducted tests.

Analysis of the data contained in the Tables1 - 3 suggests inference of the next conclusions:

a) The 10-bit quantization-based SKD system outperforms the 5-bit quantization-based one by a large margin. For the AD protocol, the average KR for

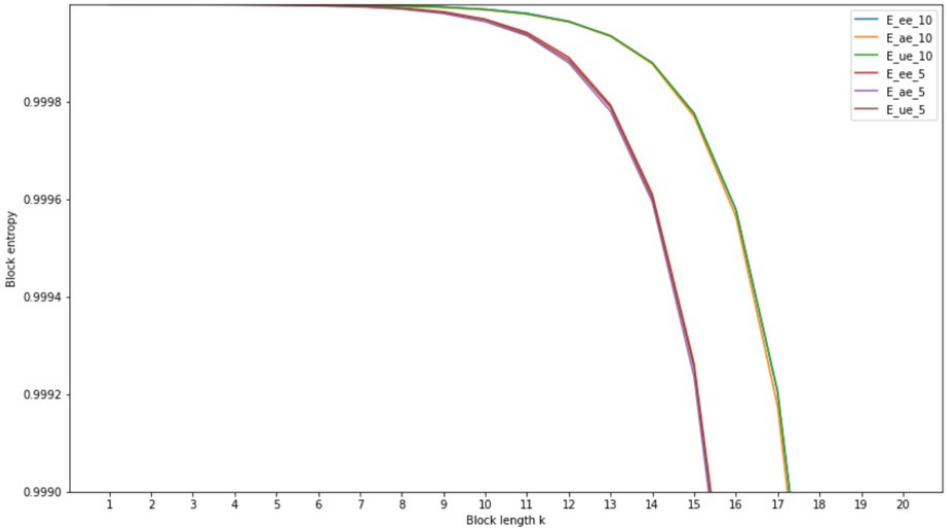

Figure 5.16: Results of the proposed SKD system regarding block entropy for booth quantization posibilities, $n_b = 5$ and $n_b = 10$ bits, for all attacker types

all Eva types is 1.78% for 5-bit quantization and 4.78% for 10-bit quantization. As a result, the 10-bit AD offers a benefit that is around 2.7 times greater than the 5-bit AD. Within the ADD protocol category, the corresponding indicators are 5.44 for 5-bit and 9.01% for 10-bit quantization; this represents a roughly 1.6-fold advantage for the 10-bit ADD.

b) For 10-bit quantization, the key agreement rate (KA) is 100%, irrespective of the kind of AD protocol

c) All tested versions of the suggested SKD system have cryptographic keys that are about equal in quality and satisfy the strictest cryptographic requirements, which incloude minimum information leakage and unpredictability (as verified by the NIST test; see Table 3). Tables 2 and 3 show that the predicted value of the NHD between Eve's and the legal keys is about 0.5, showing substantial statistical independence.

d) .A rise in communication complexity is the price paid for the high KR achieved by the ADD protocol: average IR efficiency = 3.62, as opposed to an average value of 1.17 for the AD protocol and 10-bit quantization.

It should be noted that the AD protocol's IR efficiency is nearly at unity, or the ideal value.

e) It's amazing that the attacker almost has no influence over the system's efficiency. Extremely little differences in performance indicators for each of the three Eve kinds (EE, ME, and UE) at the specified quantization and AD settings demonstrate this. The AD protocols search for signals from Bob and Alice that have a tendency to be more similar to one another than to Eve's signals, which explains this phenomena. Patterns in the principal EEG sample are clearly defined and invariant to individual differences, which appears to be how the AD methods locate these portions. The high KA rate of asynchronous EEG signals can also be explained by this process.

5.2.4 Comparison with related works

Since the SKA problem has not yet been solved using the EEG data, a direct comparison with existing research is not feasible. The most common application of EEG in the security domain is as a biometric signal for the concurrent production of cryptographic keys, which become accessible following successful authentication [65]. While it is true that these systems cannot be compared to the proposed SKA system due to the absence of the secret key delivery procedure, it is noteworthy that the most advanced systems in this class are capable of generating keys up to 192 bits, with FAR / FRR values of (0.18% / 0.18 %), [66].

One can draw indirect analogies between those works that propose an SKA based on a source model, or source of randomness, obtained from various biosignals whose sensors share capabilities similar to those of the EEG sensor. The Walkie-Talkie system is described in [67] and [68]. By taking advantage of users' gaits, or walking characteristics, two authorized devices can create a shared cryptographic key through a shared secret key generation process. It makes sense that while a person is walking, sensors at various points on the same body will sense identical accelerometer signals. According to experimental findings, the keys produced by two separate devices on the same body can reach up to 26 b/s, which calls for walking for roughly 5 seconds. Since the secret keys are established in the same physical location (the subject's body), we should be aware that this result cannot be compared to the performance of our system. This is because the secret keys are not

distributed, just like in the previously mentioned class of biometric EEG systems.

The system detailed in [69], which operates without an attacker (Eve) and utilizes the ECG signal as the source of shared randomness, is the closest conceptually to our approach. Although the authors do not verify the final keys for randomness, they report empirical results indicating a secret key creation speed of approximately $2b/s$.

In comparison to any published system within the same class, our proposed Secret Key Agreement (SKA) system, based on participants' asynchronous EEG signals, surpasses in every aspect (key generation speed, likelihood of successful key agreement, cryptographic quality of established keys, and communication efficiency).

5.2.5 Security issues and application

The suggested Secret Key Agreement (SKA) system is built upon a three-step Secret Key Distillation (SKD) procedure, demonstrating information-theoretic security. Consequently, the resulting key K, with a length of k bits, achieves maximal unpredictability, $H(K) = k$ bits, surpassing typical mathematical puzzle resolutions. Empirical determination of the final NHD (5.15) reveals its proximity to the optimal value of 0.5 (refer to Tables 1 and 2), rendering the formation of correlating keys impossible and thwarting related keys cryptanalytic attacks efficiently.

The offline and asynchronous nature of the system allows for key generation at users' convenience, thanks to the separation of the cryptographic and communications module from the source of common randomness and the SKA system. EEG signal recording can be conducted in a secure environment with varying security levels, ranging from a professional Faraday cage within a secure area to on-the-spot setups in the field. Any authenticated channel (such as the Internet) can serve as the public authenticated channel for executing the Advantage Distillation (AD) and Information Reconciliation (IR) phases of the SKD protocol.

Given the offline and asynchronous operation, the secret key agreement rate is not a critical consideration. Two use cases demonstrate the utility of the SKD system with a secret key rate of $9b/s$:

Example 4:

Assignment:: Transfer one printed page entirely covertly, ensuring absolute security based on Information theory.

Solution: Assuming an average of 20,000 bits in a printed page, the Vernam cipher requires a one-time cryptographic key of identical length. With a secret key rate of $9b/s$, a 37-minute EEG recording session is needed before generating and communicating the ciphertext.

Example 5:

Assignment: Deploy two cryptographic devices with a symmetric key of $500-$bit key length for a symmetric encryption algorithm.

Solution: Requiring an EEG signal captured for 56 seconds (500/9), or approximately one minute, both communication parties must record an EEG signal for one minute before initiating a secure connection. It's worth noting that a well-designed symmetric encryption system with a 500-bit secret key can function securely for an extended period without necessitating key changes.

5.2.6 Possibility for secret key rate improvement

To enhance the Key Rate (KR), various approaches can be considered depending on the objectives of the entire cipher system.

Approach A - Hybrid System -combination of Source and Channel Model

Following the completion of the offline procedure for secret key agreement with the proposed Secret Key Distillation (SKD) system, encrypted communication on the primary communication channel initiates. If an additional SKD based on the Channel model (SKA_ChMod) is introduced, the equivalent KR experiences a significant increase, taking into account typical KR values for SKA_ChMod systems (refer to the channel models overview, [49]). This strategy proves particularly effective for wireless main channels. However, a drawback of this approach is the vulnerability of the SKA_ChMod procedure to electronic jamming, which could lead to complete failure in critical situations such as wartime actions.

Approach B - Change of principal EEG sample

In this scenario, the Secret Key Agreement (SKA) system remains in offline mode, retaining favorable properties like robustness and high reliability. Since KR is constrained by secrecy capacity, as defined by equation (5.1), increasing it is achievable by altering the source of common randomness, aiming for a maximum Ck. In our context, this involves discovering new transformations of the original EEG signals, resulting in a primary source with a higher Ck. Additionally, within this approach, the incorporation of new biometric sensors as sources of common randomness (e.g., ECG, gait sensors) is

possible, provided it does not compromise the overall system functionality.

Approach C - Elimination of Eavesdroppers

When Eve is independent of both Alice and Bob, i.e., when Z is independent of X and Y, the secret key capacity reaches its maximum value (5.2). This scenario can be interpreted as a form of eliminating Eve, potentially creating opportunities for increased KR.

Here is a practical example of a scenario for eliminating Eve: Suppose the primary source for legitimate users Alice and Bob is established based on the transformation F_i from the set of transformations F, where any two components of this set produce outputs that are not correlated with each other (orthogonal). More specifically,

$$\begin{aligned} I\left(F_i\left(X\right), F_i\left(Y\right)\right) &> 0 \quad \forall i \\ I\left(F_i\left(X\right), F_j\left(Z\right)\right) &= 0 \quad \forall j \neq i \\ I\left(F_i\left(Y\right), F_j\left(Z\right)\right) &= 0 \quad \forall j \neq i \end{aligned}$$

for given X and Y. If denote with $|F|$ number of elements in the set F, Eva's string Z with probability $\frac{|F|-1}{|F|}$are independent of X and Y if Bob and Alice chose the transformation in secret (for example, using previously shared secret keys). In the case of deep neural networks producing transformations with millions of continuous parameters, for instance, this probability is equivalent to 1 in both theory and practice.

5.2.7 Conclusion

The study introduces a type of SKD systems whose inputs are so-called performance metrics that are obtained from communication parties' asynchronously recorded EEG signals. A careful selection of system settings can result in a key agreement rate of 100%, a secret key rate of up to 9%, good random characteristics, and a minimal information leakage (LR = 0.0003) to a possible attacker on the system, according to experimental evaluation. The system's limited sensitivity to changes in Eve's (the attacker's) EEG signal supports the theory that the synchrony of authorized participants—achieved through effective AD protocols—is crucial.

Subsequent research endeavors will center on diminishing the intricacy of communication within the suggested framework and exploring its potential amalgamation with alternative methodologies for the retrieval and dissemination of cryptographic keys, chiefly grounded in the Data Exchange

predicament [43]. Additionally, we aim to enhance the system's performance within the spheres of local randomness generation [70] and biometric applications [71]]

Chapter 6

Summary, contributions of the research work and further research

The dissertation is based on the synthesis of the publically known facts and achievements of well-known scientists with additional research on non-maximal entropy stochastic processes usage regarding randomness extraction from them.

The proposed approach introduces a novel approach for common random string establishment between communicating participants. Method offers number of benefits for application in information security solutions. Beside the direct benefits for Information security additional benefits lay in the field of continuous auditing of security in cloud computing environment. The benefit is reflected in the fact that one such method, formally based on the arguable characteristics, enables an audit based on formal and automated procedures.

The main contribution of this dissertation in the field of symmetric key establishment protocols is the following:

- **Scientific contributions:**
  - Synthesis of the Information theory methods and correlated individual biometrical signals into the novel secure protocol for symmetric secret key establishment.

- **Practical contributions:**
  - Defined protocol allows key establishment for the symmetric cryptographic systems directly between participants in communication.
  - In the cloud computing environment improve security and reduce the system complexity by elimination significant part of the key management system performed by the trusted third party.
- **Social contributions:**
  - Providing highly secure end-to-end communication in cloud computing security improve general security in the cyberspace.
  - Allow outsourced auditing of the security in the information systems.
  - Reduce complexity and cost of the information system security auditing.

Future research, as mentioned earlier, can be conducted in the direction of identification and analysis stochastic processes appropriate for this type of application especially human biometric features such as voice and eye movement.

Bibliography

[1] N. K. Sehgal, P. C. P. Bhatt, and J. M. Acken, *Cloud Computing with Security and Scalability.: Concepts and Practices*. Springer International Publishing, 2023.

[2] T. A. Kumar, T. S. A. Samuel, R. D. J. Samuel, and M. Niranjanamurthy, eds., *Privacy and security challenges in cloud computing*. Cognitive approaches in cloud and edge computing, Boca Raton, FL: CRC Press, 2022. Includes bibliographical references and index. - Description based on print version record.

[3] M. Stamp, *Information security*. Hoboken, NJ: Wiley, third edition ed., 2022. Literaturverzeichnis: Seite 409-417.

[4] J. von zur Gathen, *CryptoSchool*. Springer Berlin Heidelberg, 2015.

[5] A. J. Menezes, *Handbook of applied cryptography*. Boca Raton: CRC Press, 1997.

[6] S. D. Galbraith, *Mathematics of Public Key Cryptography*. Cambridge University Press, Mar. 2012.

[7] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, jul 1948. Available at: https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

[8] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 623–656, oct 1948. Available at: https://doi.org/10.1002/j.1538-7305.1948.tb00917.x.

[9] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. University of Illinois Press, Oct. 1963.

[10] A. D. Wyner, "The wire-tap channel," *The Bell System Technical Journal*, vol. 54, pp. 1355–1387, oct 1975. Available at: https://doi.org/10.1002/j.1538-7305.1975.tb02040.x.

[11] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Transactions on Information Theory*, vol. 39, pp. 733–742, may 1993. Available at: https://doi.org/10.1109/18.256484.

[12] R. Ahlswede and I. Csiszar, "Common randomness in information theory and cryptography. i. secret sharing," *IEEE Transactions on Information Theory*, vol. 39, pp. 1121–1132, jul 1993.

[13] M. Galis, M. Milosavljević, A. Jevremović, Z. Banjac, A. Makarov, and J. Radomirović, "Secret-key agreement by asynchronous EEG over authenticated public channels," *Entropy*, vol. 23, p. 1327, oct 2021. Available at: https://doi.org/10.3390/e23101327.

[14] M. Galis, T. Unkašević, Z. Banjac, and M. Milosavljević, "Protocols for symmetric secret key establishment: Modern approach," *Vojnotehnicki glasnik*, vol. 70, no. 3, pp. 604–635, 2022.

[15] M. Bloch and J. Barros, *Physical-Layer Security*. Cambridge University Press, 2011. Available at: https://doi.org/10.1017/cbo9780511977985.

[16] E. Y. Z. Tan, C. C. W. Lim, and R. Renner, "Advantage distillation for device-independent quantum key distribution," *Physical Review Letters*, vol. 124, Mar. 2020. Available at: https://doi.org/10.1103/PhysRevLett.124.020502.

[17] Q. Wang, X. Wang, Q. Lv, X. Ye, Y. Luo, and L. You, "Analysis of the information theoretically secret key agreement by public discussion," *Security and Communication Networks*, vol. 8, pp. 2507–2523, jan 2015. Available at: https://doi.org/10.1002/sec.1192.

[18] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin, "Experimental quantum cryptography," *Journal of Cryptology*, vol. 5, pp. 3–28, jan 1992. Available at: https://doi.org/10.1007/bf00191318.

[19] G. Brassard and L. Salvail, "Secret-key reconciliation by public discussion," in *Helleseth, T. (Eds.) Advances in Cryptology - EUROCRYPT*

'93, vol. 765, pp.410–423, Springer Berlin Heidelberg, 1992. Available at: https://doi.org/10.1007/3-540-48285-735.

[20] C. H. Bennett, G. Brassard, and J.-M. Robert, "Privacy amplification by public discussion," *SIAM Journal on Computing*, vol. 17, pp. 210–229, apr 1988. Available at: https://doi.org/10.1137/0217014.

[21] C. Cachin and U. Maurer, "Unconditional security against memory-bounded adversaries," in *Kaliski, B.S. (Eds.) Advances in Cryptology - CRYPTO '97*, vol. 1294, pp.292-306, Springer Berlin Heidelberg, 1997. Available at:
https://doi.org/10.1007/bfb0052243.

[22] A. Carleial and M. Hellman, "A note on wyner's wiretap channel (corresp.)," *IEEE Transactions on Information Theory*, vol. 23, pp. 387–390, may 1977. Available at:
https://doi.org/10.1109/tit.1977.1055721.

[23] I. Csiszar and J. Korner, "Broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 24, pp. 339–348, may 1978. Available at:
https://doi.org/10.1109/tit.1978.1055892.

[24] K. Yamazaki and T. Sugimoto, "On secret reconciliation protocol - modification of cascade protocol," in *International Symposium on Information Theory and Its applications*, Honolulu, Hawaii, pp.223–226, Nov. 5-8., 2000.

[25] T. Sugimoto and K. Yamazaki, "A study on secret key reconciliation protocol cascade," *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E83-A, no. 10, pp. 1987–1991, 2000.

[26] W. T. Buttler, S. K. Lamoreaux, J. R. Torgerson, G. H. Nickel, C. H. Donahue, and C. G. Peterson, "Fast, efficient error reconciliation for quantum cryptography," *Physical Review A*, vol. 67, p. 052303, may 2003. Available at:
https://doi.org/10.1103/physreva.67.052303.

[27] T. K. Moon, *Error correction coding*. Hoboken, N.J: Wiley-Interscience, 2005. Includes bibliographical references and index.

[28] M. Mehic, M. Niemiec, H. Siljak, and M. Voznak, "Error reconciliation in quantum key distribution protocols," in *Ulidowski, I., Lanese, I., Schultz, U., Ferreira, C. (Eds.) Reversible Computation: Extending Horizons of Computing. RC 2020. Lecture Notes in Computer Science*, pp. 222–236. Springer International Publishing, 12070, 2020. Available at: https://doi.org/10.1007/978-3-030-47361-7_ 11.

[29] R. Gallager, "Low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 8, pp. 21–28, jan 1962. Available at: https://doi.org/10.1109/tit.1962.1057683.

[30] C. Elliott, A. Colvin, D. Pearson, O. Pikalo, J. Schlafer, and H. Yeh, "Current status of the darpa quantum network (invited paper)," in *Donkor, E.J., Pirich, A.R. and Brandt, H.E. (Eds.) Proceedings Volume 5815, Quantum Information and Computation III, Defense and Security*, Orlando, Fl, March 28 - April 1, 2005. Available at: https://doi.org/10.1117/12.606489.

[31] M. Niemiec, "Error correction in quantum cryptography based on artificial neural networks," *Quantum Information Processing*, vol. 18, apr 2019. Available at: https://doi.org/10.1007/s11128-019-2296-4.

[32] M. Bloch, *Physical-Layer Security*. Cambridge University Press, Sept. 2016.

[33] P. Gronberg, "Key reconciliation in quantum key distribution," tech. rep., FOI-Swedish Defence Research Agency, 2005.

[34] S. Wolf, *Unconditional Security in Cryptography*, pp. 217–250. Springer Berlin Heidelberg, 1999.

[35] I. Csiszar and P. Narayan, "Secrecy capacities for multiple terminals," *IEEE Transactions on Information Theory*, vol. 50, pp. 3047–3061, Dec. 2004.

[36] Emotive, "Brain data measuring hardware and software solutions.," *http://emotiv.com*, 2021.

[37] Emotive, "Brain data measuring hardware and software solutions," *https://www.emotiv.com/the-science/*, 2021.

[38] J. Pourbemany, Y. Zhu, and R. Bettati, "A survey of wearable devices pairing based on biometric signals," *IEEE Access*, vol. 11, pp. 26070–26085, 2023.

[39] A. Bonci, S. Fiori, H. Higashi, T. Tanaka, and F. Verdini, "An introductory tutorial on brain–computer interfaces and their applications," *Electronics*, vol. 10, p. 560, Feb. 2021.

[40] Wikipedia, "Wisconsin card sorting test," *https://en.wikipedia.org/wiki/Wisconsin_Card_Sorting_Test*, 2021.

[41] C. A. Riccio, J. Hall, A. Morgan, G. W. Hynd, J. J. Gonzalez, and R. M. Marshall, "Executive function and the wisconsin card sorting test: Relationship with behavioral ratings and cognitive ability," *Developmental Neuropsychology*, vol. 10, pp. 215–229, Jan. 1994.

[42] M. Milosavljević, S. Adamović, A. Jevremovic, and M. Antonijevic, "Secret key agreement by public discussion from eeg signals of participants," in *5th International Conference IcEtran 2018*, (Palić, Serbia, June 11-14), 11 2018.

[43] N. Milosavljevic, S. Pawar, S. El Rouayheb, M. Gastpar, and K. Ramchandran, "Efficient algorithms for the data exchange problem," *IEEE Transactions on Information Theory*, vol. 62, pp. 1878–1896, Apr. 2016.

[44] T. A. Courtade and R. D. Wesel, "Coded cooperative data exchange in multihop networks," *IEEE Transactions on Information Theory*, vol. 60, pp. 1136–1158, Feb. 2014.

[45] A. Jevremovic, S. Arsić, M. Antonijevic, A. Ioannou, and N. Garcia, "Human-computer interaction monitoring and analytics platform - wisconsin card sorting test application," 11 2018.

[46] D. S. Benitez, S. Toscano, and A. Silva, "On the use of the emotiv epoc neuroheadset as a low cost alternative for eeg signal acquisition," in *2016 IEEE Colombian Conference on Communications and Computing (COLCOM)*, IEEE, Apr. 2016.

[47] K. Everson, "A framework for feedback control of stress using eeg and audio. bachelor's thesis," *The Ohio State University,Columbus, OH, USA,*, 2018.

[48] X. S. Zhou, L. Song, and Y. Zhang, eds., *Physical layer security in wireless communications*. No. 20 in Wireless networks and mobile communications, Boca Raton, [Florida]: Taylor & Francis/CRC Press, online-ausg. ed., 2014. Includes bibliographical references. - Description based on print version record.

[49] J. Zhang, T. Q. Duong, A. Marshall, and R. Woods, "Key generation from wireless channels: A review," *IEEE Access*, vol. 4, pp. 614–626, 2016.

[50] S. WATANABE and Y. OOHAMA, "Secret key agreement from correlated gaussian sources by rate limited public communication," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E93-A, no. 11, pp. 1976–1983, 2010.

[51] S. Watanabe and Y. Oohama, "Secret key agreement from vector gaussian sources by rate limited public communication," *IEEE Transactions on Information Forensics and Security*, vol. 6, pp. 541–550, Sept. 2011.

[52] C. Bennett, G. Brassard, C. Crepeau, and U. Maurer, "Generalized privacy amplification," *IEEE Transactions on Information Theory*, vol. 41, no. 6, pp. 1915–1923, 1995.

[53] R. A. Chou and M. R. Bloch, "Separation of reliability and secrecy in rate-limited secret-key generation," *IEEE Transactions on Information Theory*, vol. 60, pp. 4941–4957, Aug. 2014.

[54] C. Ye, S. Mathur, A. Reznik, Y. Shah, W. Trappe, and N. B. Mandayam, "Information-theoretically secret key generation for fading wireless channels," *IEEE Transactions on Information Forensics and Security*, vol. 5, pp. 240–254, June 2010.

[55] Wikipedia, "Mileševa monestary," 2021.

[56] Ward, "sklearn.cluster.ward – scikit-learn 0.15-git documentation." https://scikit-learn.org/0.15/modules/generated/sklearn.cluster.Ward.html, 2021.

[57] A. Kerckhoffs, "La cryptographie militaire.," *Journal des sciences militaires*, vol. IX, p. 5–83, Jan. 1883.

[58] A. Kerckhoffs, "La cryptographie militaire.," *Journal des sciences militaires*, vol. IX, p. 161–191., Feb. 1883.

[59] M. Gander and U. Maurer, "On the secret-key rate of binary random variables," in *Proceedings of 1994 IEEE International Symposium on Information Theory*, ISIT-94, IEEE.

[60] D. Elkouss, A. Leverrier, R. Alleaume, and J. J. Boutros, "Efficient reconciliation protocol for discrete-variable quantum key distribution," in *IEEE International Symposium on Information Theory*, Seoul, South Korea, pp.1879-1883, June 28-July 3, 2009. Available at: https://doi.org/10.1109/isit.2009.5205475.

[61] R. A., "Quantum key distribution post processing - a study on the information reconciliation cascade protocol," Master's thesis, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal, July 2019.

[62] R. Impagliazzo, L. A. Levin, and M. Luby, "Pseudo-random generation from one-way functions," in *Proceedings of the twenty-first annual ACM symposium on Theory of computing - STOC '89*, STOC '89, ACM Press, 1989.

[63] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, pp. 471–480, July 1973.

[64] L. E. Bassham, A. L. Rukhin, J. Soto, J. R. Nechvatal, M. E. Smid, E. B. Barker, S. D. Leigh, M. Levenson, M. Vangel, D. L. Banks, N. A. Heckert, J. F. Dray, and S. Vo, *A statistical test suite for random and pseudorandom number generators for cryptographic applications*. National Institute of Standards and Technology, 2010.

[65] R. Damaševičius, R. Maskeliūnas, E. Kazanavičius, and M. Woźniak, "Combining cryptography with eeg biometrics," *Computational Intelligence and Neuroscience*, vol. 2018, pp. 1–11, 2018.

[66] D. Nguyen, D. Tran, D. Sharma, and W. Ma, "On the study of eeg-based cryptographic key generation," *Procedia Computer Science*, vol. 112, pp. 936–945, 2017.

[67] W. Xu, G. Revadigar, C. Luo, N. Bergmann, and W. Hu, "Walkie-talkie: Motion-assisted automatic key generation for secure on-body device communication," in *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, IEEE, Apr. 2016.

[68] W. Xu, C. Javali, G. Revadigar, C. Luo, N. Bergmann, and W. Hu, "Gait-key: A gait-based shared secret key generation protocol for wearable devices," *ACM Transactions on Sensor Networks*, vol. 13, pp. 1–27, Jan. 2017.

[69] A. V. Guglielmi, A. Muraro, G. Cisotto, and N. Laurenti, "Information theoretic key agreement protocol based on ecg signals," May 2021.

[70] J. A. Milosavljević M., Adamović S., "Secret keys generation from mouse and eye tracking signals," in *In Proceedings of 6th International Conference on Electrical, Electronic and Computing Engineering, IcETRAN*, (Silver Lake, Serbia), pp. 1065–1068, 2019.

[71] S. Adamovic, M. Milosavljevic, M. Veinovic, M. Sarac, and A. Jevremovic, "Fuzzy commitment scheme for generation of cryptographic keys based on iris biometrics," *IET Biometrics*, vol. 6, pp. 89–96, Nov. 2016.