1 Summary

Title of the consortium research project: Cloud-assisted Security Services (CloSer)

Participants:
Industry partners:

1. F-Secure (PI: Alexey Kirichenko)
2. Nokia (PI: Yoan Miche)
3. SSH (PI: Vesa Luukkala)
4. Trustonic (PI: Jan-Erik Ekberg)

Academic partners:

1. Aalto University (PI: N. Asokan)
2. University of Helsinki (PI: Valtteri Niemi)
3. Arcada University of Applied Sciences (PI: Kaj-Mikael Björk)

2 Rationale

Background: The trend of moving previously local services to a cloud setting has become pervasive and pronounced. Security services are no exception: cloud-assisted variants of anti-malware, intrusion detection, secure storage, and application/website reputation tools and services are already being deployed. However, naïve adaptation of services to a cloud setting can give rise to new security and privacy problems. The goal of this proposal is to show how we can build effective cloud-assisted security services for organizations and individuals by designing novel security, privacy and data analytics techniques..

***Link to previous work by PIs and teams*:** The PIs and their research teams have worked together on the Cloud Security Services (CloSe) project funded by Tekes and Academy of Finland under the first ICT-2023 thematic call. This current proposal will build on the foundation laid by CloSe. The research domain at large remains the same: how to address the new security/privacy issues that arise when security services are provided from a cloud-setting and how to identify opportunities for utilizing assistance from cloud-deployed components for the purposes of security and privacy. The 2-year CloSe project had six work packages: (1) Computationally private information retrieval, (2) Secure Cloud Storage and deduplication, (3) Detecting covert/illicit channels, (4) Traffic differentiation, (5) Networking and scalability and (6) Unified design. The main thrust of CloSe was to identify security/privacy challenges in delivering "security through the cloud", especially in the context of using a "Secure Intermediary", a cloud-based secure middlebox that helps ensure security for end users. Several work packages have produced prototypes and top-level publications. For example, in WP1 we developed efficient solutions for private keyword search (PKS) using advanced cryptographic techniques. PKS enables cloud-based services that allow client devices to vet applications and documents before use without sacrificing user privacy. The solution was published at IEEE TrustCom [5] and is currently being tested by an industrial partner. In WP2, we developed a secure deduplication mechanism that is more secure and more efficient than the state of the art [1]. This work was presented at a prestigious academic security research conference, **ACM CCS** [3]. We also developed OmniShare, an application that allows users to encrypt their cloud data on client devices using strong keys while still allowing them to access the data from multiple devices. OmniShare is publicly available[1], At **CeBIT 2016**, it was awarded the **first prize** in a

[1] https://ssg.aalto.fi/projects/omnishare/

Europe-wide competition for privacy-enhancing mobile apps[2]. In WP4, we developed an efficient classification technique to identify Android malware (also published at IEEE TrustCom [4]). The Scientific Advisory Board (SAB) of CloSe consisting of world leaders in the domain[3] in reviewing progress after 10 months said [1] "CLOSE's academic progress in just one year is quite impressive" and that they were "very encouraged to witness tight academia/industry collaboration." The SAB also recognized that the ambitious doctoral research work begun in CloSe will require "3-4 years to finish".
In Section 3 we describe how we plan to build on the progress of CloSe in proposing CloSer with a refined approach and sharper research objectives.

Added value generated by the Consortium: The consortium consists of four industrial partners (as detailed in Section 7.2) and three academic partners. The industrial partners have products, which are directly in the field of cybersecurity or where cybersecurity is an essential component. The academic partners have strong track records in information security, privacy, and data analytics. We will continue the successful track record of collaboration consisting of joint publications, prototypes and technology transfer.
All partners have distinct expertise/specialization profiles in cybersecurity. Nokia is a leading manufacturer of mobile network infrastructure equipment and software. F-Secure is a provider of such security and privacy services as anti-malware, web content filtering, mobile privacy and anonymity, Advanced Threat Protection, and Smart Home protection. SSH is a software provider for trust relation management and secure communication. Trustonic is an application protection company, and its main product is a trusted execution ecosystem for mobile and embedded devices.
Each academic partner also has its own specialty – e.g., system security (Aalto), networking (UH) and technology transfer & analytics research (Arcada). Some areas of academic research expertise are shared, allowing effective joint work -- e.g., Aalto & UH (applied cryptography) and Arcada & Aalto (machine learning). These complementary as well as common expertise areas make the consortium uniquely positioned to tackle the chosen research questions.

***Collaboration in CloSer*:** The emphasis of CloSer is on developing generic enablers for combining high quality of customer protection with privacy and confidentiality of customers' data in cloud-assisted security services. CloSer will build on the collaborations set up during CloSe. In particular, close collaboration between academic and industrial partners (Section 7.2) ranges from joint problem definition to active collaborative research and **cross-sectoral mobility** between industry and academia. Our **international collaboration (**Section 7.3) with world-class research groups in information security/privacy, data analytics and networking also incorporates mobility plans for young researchers in both directions.

[2] http://mapping-competition.uni-hannover.de/winners.html
[3] Prof. Gene Tsudik (UC Irvine), Prof. Jon Crowcroft (Cambridge), Prof. Ivan Martinovic (Oxford), Dr. Ersin Uzun (PARC), Prof. Amaury Lendasse (U Iowa)

3 Objectives and expected results

By "**cloud-assisted security service**", we refer to services with an architecture where some of the service logic is implemented as a cloud service. We begin with the following core observations about cloud-assisted services and the contexts in which they operate:

1. As we saw in CloSe, moving services to the cloud raises new security/privacy challenges but straight-forward solutions may lead to tensions between conflicting requirements (e.g., security vs. deployability or security vs. usability). In CloSe, we primarily focused on two specific services: secure cloud storage and cloud-assisted malware detection. It is evident that the types of challenges as well as the nature of the conflicts they induce depend on the *particular service* under consideration. It is also evident that new or improved cloud-assisted security services are *useful and valuable in their own right*.
2. In CloSe, we have been looking at one particular instance of the secure middlebox pattern represented by F-Secure Freedome[4] which resides half-way between clients and the services they consume. But the secure middlebox pattern is more general and can be placed at different points in the continuum between clients and services. In particular, in an Internet-of-Things (IoT) setting, a secure middlebox that is close to the client devices (such as in a local home gateway like F-Secure Sense[5]) is likely to be very useful.
3. A major change is happening in the telco industry: the migration from traditional network structures to virtual systems running in a cloud. This is backed and driven by current standardization efforts, e.g. ETSI Network Functions Virtualization (NFV) Industry Specification Group. In addition, Software Defined Networking (SDN) is getting traction, as it could potentially be a significant part of this infrastructure change.

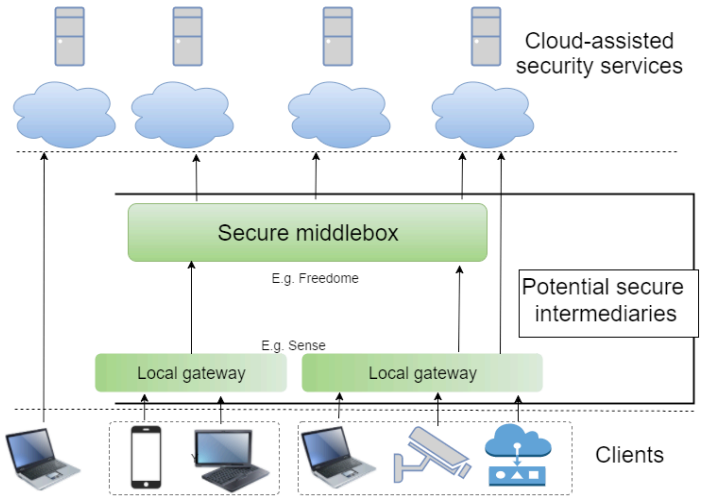

Figure 1: Cloud-assisted Security Services

Figure 1 illustrates a general model for cloud-assisted security services. People (using personal devices) and "things" will access/use various cloud-hosted security services. One or more secure intermediaries, ranging from local proxies near client devices to middleboxes serving large collections of clients, may mediate such access in order to provide better security/privacy guarantees for the clients.

[4] https://www.f-secure.com/Freedome
[5] https://sense.f-secure.com/

3.1 Research Questions and Hypothesis

The observations above raise the following questions:

1. *Cloud-assisted Security Service Scenarios*: Which usage scenarios constitute compelling cloud-assisted security services? Which of them have particular novel security and privacy concerns or tensions between security/privacy and other criteria?
2. *Privacy Enhancing Technologies:* What privacy risks do cloud-assisted security services introduce? What set of practical privacy-enhancing technologies can address them?
3. *Data Analytics Technologies:* How to devise machine learning frameworks and models that offer efficient performance under different sets of constraints? How can we utilize capabilities on the device itself or in local secure intermediaries to offer privacy-enhancing backend analytics?
4. *Infrastructure for and Integration of Cloud-assisted Security Services:* How to manage and operate software defined network (SDN)-based security middlebox networks at large scales in real-time and support distributed traffic analysis with local and distributed components? How to combine endpoint, middlebox and cloud components for a unified security solution that provides security, usability, scalability and deployability? How can trusted computing techniques be used effectively with cloud-assisted services?

Our overall **hypothesis** is *that it is feasible to design a family of effective techniques answering questions 2-4 that make it possible to develop scalable, high-performance, privacy-preserving cloud-assisted security services*.

3.2 State of the Art

3.2.1 Cloud-assisted Security Service scenarios

We consider current/potential user-facing services that raise security/privacy implications. For example:

Cloud storage servers need to authenticate clients in order to enforce policies, e.g., storage size quotas. Deduplication is widely used in cloud storage. *All* current secure deduplication schemes [2], including the one we developed in CloSe [3], have one drawback: by definition, the storage server learns which users have the same file, even if the file is encrypted. A storage client colluding with the server can thus identify all users who upload a given file.

User-tracking is a major concern in communication networks and web services. In particular, even though current (3G, 4G) mobile communication systems incorporate mutual authentication between user devices and the network, some information needs to be exchanged before authentication begins; an active attacker can set up "fake base-stations" that can track users' identities and location as we demonstrated in CloSe [6]. How to effectively avoid fake base-stations is an open research problem. A cloud-assisted service that aids users to detect/avoid user-tracking is desirable.

Application and website reputation services are used widely to steer users away from malicious or inappropriate content and functionality. Current techniques are limited in their coverage. More sophisticated solutions are needed to react to new threats without delay [18].

Cloud-hosted databases: In current cloud-hosted databases, queries that users perform as well as relational information contained in the database can both be observed by the database owner/administrator. Merely encrypting data implies that an authorized user has to download, decrypt and operate on the data locally, reducing the value of a cloud-assisted service. Attempts have been made to allow processing of encrypted databases directly [13], [14] with limited success [15]. Trust relation management systems, such as SSH's Universal Key Manager[6] maintain a representation of managed networks which can be abstracted as graphs

[6] http://www.ssh.com/products/universal-ssh-key-manager

where each node represents a subject or object and edges represent access control conditions (such as the fingerprint of a key authorized to access an object). Storage and operations on such graphs can be realized by a database, but in order to be truly privacy preserving, such a database should support queries of the form "Can X access Y?" revealing (a) no information about the query to the database owner/administrator and (b) no additional information to the querier about the graph except the answer to the query itself (e.g., whether a path from X to Y exists in the graph without learning what the intermediate nodes are). Although there has been some work on making queries on a graph in a privacy-preserving manner [16][17], there is no comprehensive design of privacy-preserving graph databases in the research literature.

3.2.2 Privacy-enhancing technologies

The best known standard cryptographic techniques for computationally private information-retrieval (e.g.,[9]) are expensive in terms of computation and communication requirements. But specific problem scenarios like privacy-preserving application reputation services may admit more efficient solutions that take advantage of particular characteristics of the scenarios (such as the ability to tolerate a moderate level of false positives). In CloSe we developed an efficient PKS solution for cloud-assisted malware detection [5] but it still requires an occasional communication cost that is linear in the size of the database and a constant number of public key operations for each query. Solutions using trusted hardware on the server [8] or garbled circuits for multiparty computation (MPC) hold the promise of significantly better performance.

3.2.3 Data analytics technologies

To maximize privacy, the bulk of the machine learning processing must take place close to the client devices. However, resource constraints on client devices limit the amount of processing on the devices themselves suggesting that a cloud-assisted approach is beneficial. Currently there is no framework for hardware- or (data) location-constrained machine learning that guarantees good accuracy vs. time tradeoff. Traditional ensemble models are insufficient in this setup as they only allow for simple weighting of individual models, not considering several possible constraints at the same time in their weighting.

Our earlier work on web content filtering through image classification found promising results, although the performance was still insufficient for practical applications, particularly due to the significant number of false positives [11]. Recently, deep learning methods with convolutional networks have shown great potential in various image analysis tasks [12].

3.2.4 Infrastructure for cloud-assisted security services

Software-defined networking (SDN) has been applied to specific types of secure intermediaries, e.g. for anomaly detection [17]. Docker technology is widely applicable to cloud computing scenarios. However, no straight-forward solutions are known for applying such technologies for arbitrary functionalities that could be included in secure intermediaries.

Standard techniques for using trusted hardware for building trust in remote services, such as attestation, assume that a client can identify and validate a single physical hardware trust anchor such as a Trusted Platform Module (TPM) embedded in a server. Capabilities such as hardware-enforced Trusted Execution Environments and sealed storage assume that the computation always takes place on the same physical hardware. However, these contradict standard cloud practices like virtualization, workload migration and elasticity (i.e. dynamically scaling an application using multiple virtual machines). A combination of new hardware security mechanisms (like Intel SGX[7]) with privacy-preserving attestation schemes [10] may offer one possible direction to resolving this contradiction. No current hardware

[7] https://software.intel.com/en-us/isa-extensions/intel-sgx

platform security mechanisms (such as TPM or SGX) or standards (such as those from GlobalPlatform[8] or ETSI NFV[9]) deal with migration of trust anchors in a cloud setting, or upscaling/downscaling of applications that use these roots of trust. Santos et al [21] describe "policy-sealed data" as a solution to allow multiple nodes (whose configurations match the stated policy) to access sealed data. However, they do not address the issue how such policy-sealed data can thwart replay attacks.

3.3 Scientific and societal impact

The expected results include high-quality scientific articles addressing the research questions, prototype implementations of the proposed techniques, unified design patterns for cloud security services, and their respective evaluation as well as technology transfer for the participating industry partners. The scientific results are expected to pave way for the next generation cloud-assisted services for which security and privacy are crucial requirements.
Concrete research results for research question 1 will be compelling demonstrators which may be exploited on their own. Results for questions 2 and 3 will be concepts, mechanisms, possible trade-offs between different goals, implementations and performance analyses. Answers to question 4 are expected to provide mechanisms for scalable distributed security functions. Finally, the unifying thesis is expected to be solved by system architecture and related concepts, mechanisms, prototype implementations and thorough analysis. By enabling more secure and privacy-preserving cloud-assisted services our overall societal impact is improving protection for people and organizations who use and benefit from such services.
In addition to solving specific problems of CloSer industrial partners and enhancing their competitiveness, CloSer will continue to directly contribute to the building and extending of cybersecurity competence and community in Finland.

3.4 Applicability of research results

The four industrial partners (F-Secure, Nokia, SSH, and Trustonic) expect that collaboration with the academic partners will provide significant value for building a solid foundation for their new solutions and services. As input to the project, they will provide requirements, field expertise, appropriate data sets and security platforms (on which research prototypes and the final design can be validated). Industrial partners plan to have their research teams work closely together as well as to invite team members of the academic partners to work in the company environments and by establishing internships. They believe that CloSer provides excellent opportunities for **solving real-life problems** by conducting high-class collaborative research in the domains of user privacy as well as data, device, and network security.
Industrial partners expect that CloSer results will contribute to their products and services. Whenever appropriate, they will publicize the fact that such services utilize approaches and technologies developed based on results from CloSer. Concrete demonstrators of new or improved cloud-assisted security services developed in addressing research question 1 may be productized by CloSer industrial partners and industry at large. New technological enablers developed in addressing research questions 2 and 3 will, in addition to being useful in demonstrators, be applicable in designing other solutions.

3.5 Critical success points and risk management

The project will be divided into three phases with identified milestones as described in Section 6.1.5. Potential risks and their mitigation are discussed in Section 6.1.6.

[8] https://www.globalplatform.org/
[9] http://www.etsi.org/technologies-clusters/technologies/nfv

3.6 Publication Plan

The results will be published in top conferences and journals, i.e. in fora like ACM CCS, IEEE S&P, Usenix Security, NDSS, ACM SIGCOMM, ACM CoNEXT, ACM ICN, IEEE/ACM Transactions on Networking, Journal of Machine Learning Research, Neurocomputing and Information Sciences. Results will be made available to the public via the project website and public demonstrations. The final results workshop will be public. Project results will be incorporated in the courses taught at academic partner institutions, e.g., the Mobile System Security course at Aalto University and the Cryptography and Network Security course at University of Helsinki.

4 Research methods and material, support from research environment

Research methods: Several different research methods are employed in the project, namely cryptographic protocol design and analysis, system design, machine learning and data analytics. The theoretical designs are experimented with prototype implementations in a laboratory environment to study their properties. Specific research methods are explained in greater detail in Section 6.1 where we describe individual work packages.

Data management plan*:* Research data will consist mainly of data collected by our industrial partners and, in some cases, openly available data sources and data sets of our collaborators will be used. The primary goals of the research data use are (a) better understanding of research work objectives, priorities, and constraints; (b) forming of initial ideas of viable research approaches and evolving those; (c) building and validating of data analysis algorithms, models, and other relevant methods and techniques.

All CloSer partners will conform to the European and national data protection and privacy regulations, based primarily on the European directive 95/46/EC, regarding the collection of data for the purposes of scientific research. The partners will avoid unnecessary collection and use of personal data (data retention) and abide by the principle of informed consent for data falling in that category.

The partners will implement appropriate IT-security and organizational measures to protect data and address privacy concerns to the best extent possible, in particular:

1) All personal data will be anonymized or pseudonymized (by replacing identifiable data portions with pseudonyms).

2) Access control rules and access logging will be established (i.e. only named researchers in the consortium are given access to specific parts of raw data on the need basis).

3) The researchers will be instructed as to what was communicated to the users, who agreed to share some of their data and activity information, and required to comply with it. In particular: no de-anonymization, only analyze for the intended purpose, delete personal data after 6 months. Permission from the users will be re-queried by email after every 6 months.

4) When required, the researchers will sign appropriate non-disclosure agreements to get access to data.

Data sets used for research and not containing personal data or other sensitive information may be shared with external parties, subject to approval of the respective data set owners. In any case, the best effort will be taken to ensure that the data protection arrangements will not impede publication of the scientific results in conferences and/or journals.

5 Ethical issues

Ethical issues involved in data collection are discussed above in the data management plan (Section 4). No additional ethical issues are foreseen.

6 Implementation: Timetable and distribution of work

We will structure the work in CloSer into four work packages with three phases.

6.1 Work Packages and Schedule

6.1.1 Work Package 1: Cloud-assisted Security Service (CloSer) Scenarios

This work package will identify selected, compelling, usage scenarios for cloud-assisted security services that (i) involve difficult privacy or data analytics challenges which can then be investigated further in the subsequent work packages and (ii) are interesting/useful in themselves. In this work package, we will **design, implement and demonstrate** solutions for selected **usage scenarios**. Initially, we identify the following example usage scenarios:

a) Secure cloud storage: Using trusted hardware, design a secure deduplication scheme that prevents even the storage server from learning which users share a file while still allowing it to enforce quotas and do deduplication.

b) Detection of user-tracking in communication networks: Design a cloud-assisted collaborative service that allows current (3G, 4G) or future (5G) mobile devices to detect and avoid fake base-stations.

c) Extending capabilities of application and website reputation services: Design new capabilities that can be used by such reputation services. As examples, we expect to develop (i) a cloud-based package profiling service for Android apps developer keys to warn users about to install a (purportedly popular) app signed with an untypical key; and (ii) novel techniques to categorize web content in order to improve the performance of web filtering services.

d) Cloud-assisted trust relation database: Design a privacy-preserving trust relation database that can answer queries of the form "Can X access Y?" without revealing any additional information to the querier or the database administrator.

Task	Name	Deliverables
1.1	Secure deduplication with trusted HW	Design and prototype
1.2	Cloud-assisted user-tracking and fake base station detector	Feasibility study, designs, one or more prototypes.
1.3	App-signing key reputation service	Design and prototype
1.4	Web site reputation service	Design and prototype
1.5	Cloud-assisted trust relation database	Design and prototype

Participants: F-Secure, Nokia, SSH, Trustonic, Aalto, UH, Arcada

6.1.2 Work Package 2: Privacy-enhancing Technologies for CloSer

Informed by the usage scenarios in WP1, we will

(a) seek to **understand how user privacy is impacted** by cloud-assisted security services: For instance, example scenario 'c)' in WP1 requires a user device to query a cloud service with the list of apps installed on the device. Prior research [7][20] has shown that this allows the cloud service to infer user characteristics (e.g., gender, age, interests, preferences etc.) but the full extent of the resulting privacy exposure is not fully understood yet. We will systematically and rigorously quantify the extent to which personal characteristics of users can be inferred by examining the set of applications installed on that user's device.

and

(b) develop a **suite of privacy-enhancing technologies** for ensuring privacy for end user and customer and organizational data in WP1 and other scenarios:

In CloSe we developed a solution for private key word search (PKS) using cryptographic techniques. In CloSer, we plan to investigate additional usage scenarios and develop new significantly more efficient and expressive PKS, including techniques leveraging trusted hardware on the cloud server side.

In addition, we will investigate how to design practical techniques for supporting privacy-preserving graph queries. Direct application of secure multiparty computation to this problem is prohibitively expensive. Our initial approach will be to identify access control conditions in the specific databases that are not privacy-sensitive, and then making these public. We can thus potentially render the problem of privacy-preserving graph queries tractable.

Task	Name	Deliverables
2.1	Privacy impacts of cloud-assisted services	Survey
2.2	Efficient private keyword search	Using trusted hardware: design and prototype; Using multiparty computation (MPC): design and prototype
2.3	Privacy-preserving database queries	Using MPC; design and prototype. Validation with real-world data from industrial partners

Participants: F-Secure, Nokia, SSH, Trustonic, Aalto, UH

6.1.3 Work Package 3: Data Analytics Technologies for CloSer

We will develop novel data analytics approaches that preserve privacy while fitting within emerging cloud-based network infrastructure paradigms such as software-defined networking (SDN) and Internet of Things (IoT). The research problem stems from the fact that data sent by IoT devices or end-points in organizations e.g., can be too sensitive to be processed as-is by a remote server in the cloud. To address privacy concerns arising in such scenarios we aim to do prediction and anomaly detection tasks at or close to client devices as much as possible within their time and hardware constraints. We will develop machine learning and data analytics tools capable of performing complex data analysis tasks in a privacy-preserving and localized manner, with guarantees over the quality of their prediction or anomaly detection capabilities. This raises the second research topic: how to design meta-models which work at the backend level, with the predictions and processed data sent to them by client devices. These models need to be able to make aggregate decisions, with varying importance of the sub-models feeding data to them.

These two research tracks will result in a distributed, privacy-aware computation framework directly supporting scenarios ‘b)’ and ‘c)’ from WP1. For example, in scenario ‘c)’, we will explore content filtering models where parts of the processing, e.g., feature extraction, are done at the endpoint, while heavy analysis and model training with advanced time-consuming methods are done in the cloud. This will require new approaches to image analysis, in particular for extracting appropriate features from images. The feature set should be compact enough for efficient storage and transmission, yet be sufficiently expressive to allow accurate categorization for the purposes of filtering. The cloud service could store a database of extracted features (not needing access to the original images), and use this for machine learning. Only transmitting and storing a limited, abstract feature set would alleviate potential privacy concerns. Adopting a suitable feature extraction scheme is a crucial step for designing the framework; privacy-preserving SIFT [21] is one promising option. A key issue will be how to evaluate the full framework, as the common approach of randomly sampled training and testing sets does not correspond to the reality of a constantly evolving threat landscape.

Task	Name	Deliverables
3.1	Identify requirements for image classification services	Survey
3.2	Improved Machine learning techniques for image analysis	New methods, algorithms, validation with data from industrial partners.
3.3	Distributed machine learning framework for IoT	Framework design, algorithms, prototypes, testing in controlled IoT environments.
3.4	Machine learning for apps and devices classification	New algorithms and models; validation with data from industrial partners.

Participants: F-Secure, Nokia, Arcada, Aalto

6.1.4 Work Package 4: Infrastructure for and Integration of CloSer

The focus of this work package is to tackle the infrastructure concerns in cloud-assisted security services. We will have two directions.

First, we will extend the current secure intermediary prototype developed in the CloSe project with local and cloud supported data analytics technologies (WP3) as well as privacy-enhancing technologies (WP2). The prototype will also be used to demonstrate some of the scenarios studied in WP1, e.g. 'c)' and d)'. The secure intermediary is SDN-based and supported by a scalable Docker-based cloud. We investigate the scalability aspects of the solution and develop a cloud-based management plane. We will also investigate the integration of endpoint security and the secure intermediary system. It is expected that even a **light-weight support from end points**, e.g. smartphone running minimal security software, will significantly improve security and performance of the overall system. The goal is real-time detection/mitigation of network threats without significant overhead or privacy risk. We will also consider scenarios where security solutions have no end-point presence.

Second, we will develop new schemes that reconcile the use of trusted hardware with standard cloud computing practices such as virtualization, workload migration and elasticity. To facilitate remote attestation and secure storage of migratable, virtualized services without having to make this migration visible to clients, we will explore how roots of trust can be virtualized and *efficiently migrated and scaled* on-demand. A starting point will be to use group signature mechanisms with attestation schemes for specific environments like TPM-2 and SGX. We will demonstrate these schemes with selected WP1 scenarios (e.g., 'b)' or 'c)') as well as make them available as enablers for secure intermediaries such as F-Secure Sense.

Task	Name	Deliverables
4.1	Extensions of the secure intermediary prototype with selected privacy enhancing mechanisms (WP2) and local/cloud-supported data analytics methods (WP3)	Integrated prototype
4.2	Development of cloud-based management plane for the secure intermediary	Design and prototype
4.3	Development of a light-weight client based security mechanism and its integration to the secure intermediary solution	Feasibility study, design
4.4	Efficient schemes for using trusted hardware to achieve protected execution, secure storage, and trustable attestation in cloud environments.	Feasibility study, design, prototype
4.5	Demonstration of selected scenarios (WP1) integrated with the secure intermediary.	Demonstrator

Participants: F-Secure, Nokia, SSH, Trustonic, UH, Aalto

6.1.5 Project schedule and milestones

Figure 2 shows interactions in CloSer. Results from WP2-4 will be useful in other contexts than cloud-assisted security services. CloSer demos from WP1 will be of independent interest as standalone services in themselves. We will have three phases as shown below:

	Duration	**Description**
1	Y1M1-Y1M6	Selection & initial design of WP1 scenarios; Formulation of WP2-4 requirements
2	Y1M6-Y2M9	Iterative design and implementation WP2-4 technologies and WP1 scenarios
3	Y2M9-Y2M12	Refinement and wrapping up

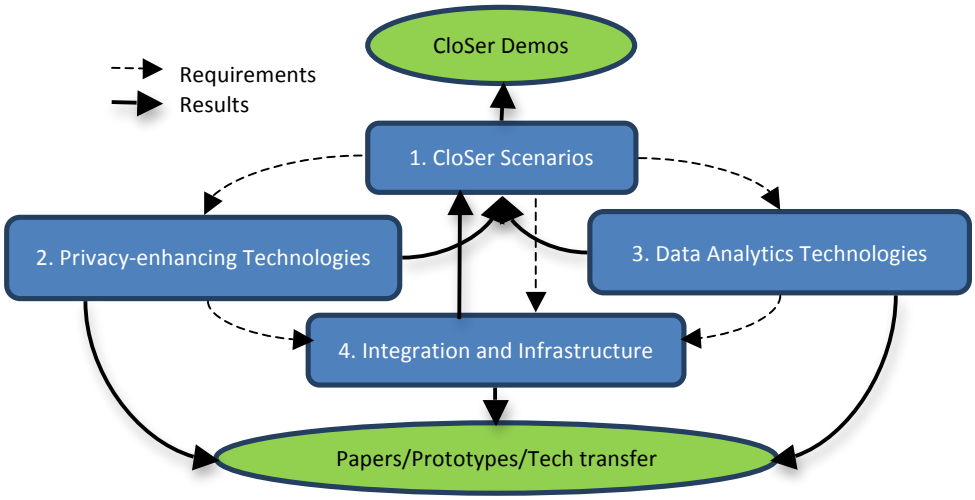

Figure 2: Structure of CloSer

6.1.6 Risk management

As in any research endeavor, there is some uncertainty associated with the assumptions and expectations in this project. The risks and the mitigation measures we identified are:

#	Risk	Task/ WP	Likelihood	Mitigation
1	Scenarios identified in 6.1.1 are not feasible	1	medium	By Y1M6 (Milestone #1) identify alternative CloSer scenarios to realize.
2	PKS schemes developed cannot handle sufficiently large databases.	2.2	low	For hardware-based PKS, use parallelization to increase the size of database supported.
3	Traditional feature extraction methods prove insufficient for the image classification task	3.2	medium	Deep learning convolutional neural networks are explored for feature extraction.
4	Datasets collected/obtained are not adequate.	3.3	high	Generate simulated data using purpose-built CloSer testbed.
5	The secure intermediary solution cannot be scaled.	4.1	low	Modify intermediary by (a) moving some functionality closer to the client or (b) sacrificing some functionality.
6	No client solution can be found that is both secure and lightweight.	4.4	medium	A solution is developed that does not assume any significant contribution from clients.
7	No cryptographic solution for supporting attestation of migratable services is found.	4.5	medium	Design a standalone attestation service that can offer property-based attestation of services hosted in the same domain.

6.2 Project meetings and governance

Each work package will meet frequently for technical discussions. Following the practice in CloSe, results will be presented in public workshops[10]. A steering group consisting of the PIs and representatives of the four industrial partners will meet during the workshops (and as needed otherwise) to evaluate project progress. The CloSer scientific advisory board (SAB) will consist of Profs Jon Crowcroft (Cambridge), Ivan Martinovic (Oxford) and Gene Tsudik (UC Irvine) and Dr. Ersin Uzun (Palo Alto Research Center). The SAB will be invited to results workshops for feedback and guidance.

7 Teams and collaboration

7.1 Academic Partners

The team from **Aalto** University ((http://cs.aalto.fi/en/secure_systems/), Department of Computer Science is led by PI **Prof. N. Asokan** and co-PI **Dr. Andrew Paverd**, and includes **one post-doctoral researcher and three PhD students**. It brings expertise in applied cryptographic protocols, system security, machine learning and security usability. Prof Asokan is well recognized for his extensive research in and contributions to system security. His research results are widely deployed and widely cited. He is an associate editor of IEEE Security & Privacy and ACM TISSEC, both leading venues for security/privacy research. He was granted a Google Research Award and was recently named ACM Distinguished Scientist.

The team from University of Helsinki (**UH**) is led by **Prof. Valtteri Niemi** and includes **one post-doctoral researcher and three PhD students**. The team has expertise in cryptography and its applications, security of mobile communications, cloud platforms, and networking. Prof. Niemi has co-authored four books on security of mobile communications, has led the standardization of 3G and 4G security in 3GPP for 6 years and has co-authored several patents that are essential for cellular standards.

The team from **Arcada** University of Applied Sciences (http://www.arcada.fi/en) is led by Head of Department Kaj-Mikael Björk and includes **two post-doctoral researchers**. Dr. Björk is heading the department of Business Management and Analytics as well as the research group in Analytics in Arcada. He is also an adjunct Professor in Åbo Akademi University (in Information Systems). The team consists of researchers in both Business Analytics and Information Analytics.

7.2 Industrial Partners

CloSer consortium has four industrial partners who represent different sectors in the Finnish security industry: a multinational corporation (Nokia), two medium sized companies (F-Secure and SSH) and the Finnish subsidiary of a small European company (Trustonic). Nokia, F-Secure and Trustonic are already partners in the CloSe consortium. They helped formulate research questions, participated in research, and are testing selected results. We expect the same pattern of successful collaboration in CloSer.

F-Secure's objective in the project is to develop novel ways of protecting organizations and individual users via security services supported by cloud-based capabilities. People and organizations are facing major threats, such as targeted attacks, malicious objects and content, on-line tracking, misbehaving devices and apps, compromised public Wi-Fi networks, and so on. Since purely end-point protection mechanisms are inadequate, and often infeasible, for solving such security problems, F-Secure is exploring ways of utilizing cloud computing for advanced data analysis and methods based on the "secure intermediary"

[10] https://wiki.aalto.fi/display/CloSeProject/CloSe+Project+Workshop

paradigm. A number of key challenges that we need to address in this work fit perfectly in the CloSer research agenda.
Nokia focuses on researching methods and systems for differentiating legitimate and malicious traffic in the project's research problem area. Their objective is to develop a semi-autonomous and privacy-enhancing monitoring system that utilizes artificial intelligence for recognizing malicious traffic and usage patterns, scaling up to global services.
Trustonic is looking for ways to exploit trusted execution environments technology in servers, i.e., in a cloud context. Also, they want to provide device and user authentication for selected services researched and developed in this project and to contribute to the developing project concepts in trusted hardware.
The objective of **SSH** is to research and develop novel solutions for analysing and operating on sensitive customer data in a privacy preserving manner, typically in a multiparty environment where parties have partial ownership of the data. One of the most central themes is trust relationship data, which is usually augmented with other business process related information.

7.3 National and international Academic Partners

We will continue our intensive and focused collaboration with our **core collaborators.** They include international core collaborators -- **Prof. Benny Pinkas** (Bar Ilan University**;** applied cryptography), **Prof. Jörg Ott** (Technical University of Munich**;** networking), **Prof. Thomas Schneider** (Technical University of Darmstadt, applied cryptography), **Prof. Ahmad-Reza Sadeghi** (Technical University of Darmstadt**;** system security) and **Prof. Amaury Lendasse** (University of Iowa; machine learning) – and national core collaborators – **Prof. Juha Karhunen** (Aalto University**;** machine learning) and **Prof. Sasu Tarkoma** (University of Helsinki**;** networking). We expect several research visits (in both directions) with international core collaborators. In addition to the core collaborators, we expect to co-operate with other academic and industrial researchers around the world on selected topics.

Partner	Description	Tasks
Prof. Benny Pinkas, Bar Ilan University	Secure cloud storage, improved PMT	Tasks 1.1, 2.2
Prof. Thomas Schneider, TU Darmstadt	PMT using MPC	Task 2.2
Prof. Jörg Ott, TU Munich	Reconciling trusted HW with cloud practices, secure intermediary scalability	Tasks 4.1-4.5
Prof. A-R Sadeghi, TU Darmstadt	Secure cloud storage	Task 1.1
Prof. Juha Karhunen, Aalto University	Machine learning specifications /improvements	Tasks 3.1, 3.2
Prof. Sasu Tarkoma, UH	Secure intermediary architecture	Tasks 4.1-4.4
Prof. Amaury Lendasse, U Iowa	Machine learning and image analysis	Task 3.2

8 Researcher training and research career

The core research team consists of 7 PhD students and 3 post-doctoral researchers. Some of the PhD students began their research already in CloSe and are expected to complete it during CloSer. Post-doctoral researchers will gain experience both in teaching and in guiding younger researchers. Instruments used in the project, such as regular workshops and the planned researcher mobility will expose researchers to inter-disciplinary and international collaboration. The PIs will be directly responsible for the supervision of the researchers. All PhD students participate in the Helsinki ICT Doctoral Education Network (HICT) network to broaden their expertise and interact with peers.
All academic and industrial partners are committed to support and value diversity and equal opportunity of individuals working in/with their respective organizations.

9 Mobility plan

***Cross-sectoral mobility**:* We plan 2-way mobility between industrial and academic partners in CloSer including **multiple internships** of academic project researchers at the industrial partner companies and short visits to university research groups by engineers in partner companies who are pursuing higher education. At least **three industry researchers** are expected to do doctoral research work in the context of CloSer. Industrial partners will make it possible for the partners to work on their premises for integrating technology prototypes to appropriate laboratory systems and validating their performance. We plan several research visits by project researchers to groups of international collaborators (see Section 7.3) as well as host visits by collaborators and other leading researchers in cloud security.

- *Outgoing mobility:*

Partner	Participant	Sub-Project	Description
Bar Ilan U	Liu	Aalto	Secure cloud storage, PMT using trusted HW & MPC (Tasks 1.1 & 2.2)
U Iowa	Akusok	Arcada	Image classification (Task 3.2)

- *Incoming mobility:*

Partner	Participant	Sub-Project	Description
Bar Ilan U	Benny Pinkas	Aalto	Secure cloud storage, PMT using trusted HW & MPC (Tasks 1.1 & 2.2)
U Iowa	A. Lendasse	Arcada	Image analysis (Task 3.2)
TU Munich	Jörg Ott	UH	Secure intermediary scalability (Tasks 4.1-4.4)
U Agder	V. Oleschuk	UH	Privacy-preserving database queries (Task 2.3)
Xidian Univ.	NN	UH	Privacy impacts of CloSer (Task 2.1)

10 Mechanisms for integration of partners

CloSer will continue all the mechanisms being used successfully to integrate partners in CloSe:

- **Mailing list, wiki, source code repositories**: CloSer will have several mailing lists (steering group, all hands, and individual WPs). CloSer wiki has a public section for dissemination of results and an internal section, accessible to all CloSer participants, for coordinating work. Each WP will maintain source code repositories as needed.
- **Regular WP meetings**: Each active WP will meet at least bimonthly to discuss progress and plan work. Each WP meeting is announced to the entire CloSer mailing list.
- **Collection and sharing of datasets**: CloSer will collect datasets needed for research. Collection, protection and use of datasets will follow the principles established in CloSe.
- **Joint demonstrations and research papers:** CloSer will hold two public workshops for showcasing research results in the form of demonstrators and posters. They will be modeled after the overwhelmingly successful first public workshop of CloSe[11], attended by over a hundred participants. In addition, CloSer results will be included in other suitable demonstration opportunities (e.g., the UH and Aalto Secure Systems groups hold an annual demo day). All employees of CloSer partners, including those who are not directly working in CloSer will be invited to these events. We also expect to continue to produce several top-tier research papers with joint university and industry authorship.

[11] First CloSe Public Workshop: https://wiki.aalto.fi/display/CloSeProject/CloSe+Project+Workshop

Bibliography

[1] CloSe Scientific Advisory Board: Report of the first CloSe public workshop, https://goo.gl/6eZQDy

[2] Mihir Bellare, Sriram Keelveedhi, and Thomas Ristenpart: DupLESS: Server aided encryption for deduplicated storage. Usenix SEC 2013. http://eprint.iacr.org/2013/429

[3] Jian Liu, N. Asokan, and Benny Pinkas: Secure Deduplication of Encrypted Data without Additional Independent Servers. ACM Computer and Communication Security (ACM CCS), 2015. (Extended version at http://eprint.iacr.org/2015/455)

[4] Luiza Sayfullina, et al: Efficient detection of zero-day Android Malware using Normalized Bernoulli Naive Bayes. 14th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (IEEE TrustCom-15), 2015 http://dx.doi.org/10.1109/Trustcom.2015.375

[5] Tommi Meskanen, et al: Private Membership Test for Bloom Filters. 14th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (IEEE TrustCom-15), 2015 http://ieeexplore.ieee.org/xpls/icp.jsp?arnumber=7345322

[6] Ravi Borgaonkar, et al: Practical Attacks Against Privacy and Availability in 4G/LTE Mobile Communication Systems. 23rd Annual Network & Distributed System Security Symposium (NDSS), February 2016, http://www.internetsociety.org/events/ndss-symposium-2016/ndss-2016-programme

[7] Suranga Seneviratne et al: Predicting user traits from a snapshot of apps installed on a smartphone. Mobile Computing and Communications Review 18(2): 1-8 (2014) http://doi.acm.org/10.1145/2636242.2636244

[8] Michael Backes et al: Oblivad: Provably Secure and Practical Online Behavioral Advertising. 2012 IEEE S&P, May 2012 http://dx.doi.org/10.1109/SP.2012.25

[9] Femi Olumofin and Ian Goldberg: Revisiting the Computational Practicality of Private Information Retrieval. Financial Cryptography 2011 http://dx.doi.org/10.1007/978-3-642-27576-0_13

[10] Ernest F. Brickell, Jan Camenisch, and Liqun Chen: Direct anonymous attestation. ACM Conference on Computer and Communications Security 2004 http://doi.acm.org/10.1145/1030083.1030103

[11] Anton Akusok, et al: Arbitrary Category Classification of Websites Based on Image Content. Computational Intelligence Magazine, IEEE, vol.10, no.2, pp.30-41, 2015 http://dx.doi.org/10.1109/MCI.2015.2405317

[12] Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep Learning. Book in preparation for MIT Press, 2016 http://www.deeplearningbook.org/

[13] Raluca Ada Popa, et al: CryptDB: Protecting Confidentiality with Encrypted Query Processing. 23rd ACM Symposium on Operating Systems Principles (SOSP), Cascais, Portugal, October 2011. http://people.csail.mit.edu/nickolai/papers/raluca-cryptdb.pdf

[14] Nathan Dowlin, et al: Manual for Using Homomorphic Encryption for Bioinformatics. Microsoft Research Techreport MSR-TR-2015-87, http://research.microsoft.com/apps/pubs/default.aspx?id=258435

[15] Muhammad Naveed, Seny Kamara, and Charles V. Wright: Inference Attacks on Property-Preserving Encrypted Databases. ACM Conference on Computer and Communications Security 2015: 644-655, http://doi.acm.org/10.1145/2810103.2813651

[16] Xianrui Meng,et al: GRECS: Graph Encryption for Approximate Shortest Distance Queries. ACM Conference on Computer and Communications Security 2015: 504-517, http://doi.acm.org/10.1145/2810103.2813672

[17] David J. Wu, et al: Privacy-Preserving Shortest Path Computation. 23rd Annual Network & Distributed System Security Symposium (NDSS), February 2016, http://www.internetsociety.org/events/ndss-symposium-2016/ndss-2016-programme

[18] Kostas Giotis, et al: Combining OpenFlow and sFlow for an effective and scalable anomaly detection and mitigation mechanism on SDN environments. Computer Networks 62: 122–136, 2014. http://www.sciencedirect.com/science/article/pii/S1389128613004003

[19] Xiaoguang Qi and Brian D. Davison. Web Page Classification: Features and Algorithms. ACM Computing Surveys, 41(2), February 2009. http://dx.doi.org/10.1145/1459352.1459357

[20] Eric Malmi and Ingmar Weber: You Are What Apps You Use: Demographic Prediction Based on User's Apps. ICWSM 2016, http://arxiv.org/abs/1603.00059

[21] Nuno Santos, et al: Policy-Sealed Data: A New Abstraction for Building Trusted Cloud Services. USENIX Security Symposium 2012: 175-188 https://www.usenix.org/conference/usenixsecurity12/technical-sessions/presentation/santos

[22] Chao-Yung Hsu, Chun-Shien Lu, and Soo-Chang Pei: Image Feature Extraction in Encrypted Domain With Privacy-Preserving SIFT. IEEE Transactions on Image Processing, vol. 21, no. 11, pp. 4593-4607, Nov. 2012. http://dx.doi.org/10.1109/TIP.2012.2204272