1 Summary

Title of the consortium research project: Cloud Security Services (CloSe)
Consortium PI and site of research: Prof. Valtteri Niemi, University of Turku
Subproject PIs and sites of research:
Profs. N. Asokan (CSE), Juha Karhunen (ICS), Jörg Ott (Comnet), all from Aalto University
Dr. Amaury Lendasse, Arcada University of Applied Sciences
Prof. Sasu Tarkoma, University of Helsinki.

2 Background

Threats for cloud services: The emerging cloud computing paradigm is changing the nature of computing and its applications in profound ways. There is a distinct trend towards hosting application and service logic in the cloud while keeping the client applications lean. In this model, user and application data are stored in the cloud and they are used on client devices in conjunction with appropriate caching strategies. The cloud paradigm naturally also enables hosting various security services with the benefit of elasticity and higher independence of the execution environment. This facilitates greater freedom in the placement of security services and their components as well as allows network-level control of the data flows.

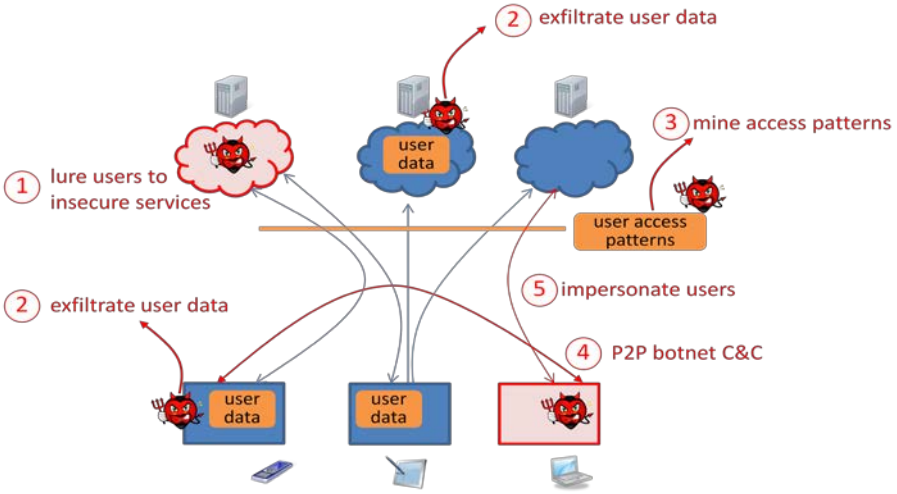

Figure 1 Threats for Cloud Services

We have identified a number of current and emerging security threats related to such usage scenarios (Figure 1):

(1) Users can be lured to insecure/malicious services or web resources.
(2) Sensitive user data can be exfiltrated from the cloud services or from user devices.
(3) Access patterns and traces of other activities of the user can be mined to reveal sensitive information about the user.
(4) Illicit activities may originate from user devices, e.g. for the purpose of running a peer-to-peer control system for a botnet. This threat may be magnified if countermeasures against (3) hide malicious operations as well.
(5) The user can be impersonated, for instance, via identity theft, to gain access to services/resources intended for the user.

These threats can be addressed by using a variety of mechanisms. For example, user data can be **encrypted** before being stored on cloud services; data access patterns may be **hidden** using series of intermediaries (as in TOR), decoy traffic, and/or decoy data (one flavor being steganography). A logical entity called **Secure Intermediary** can steer users away from

insecure/malicious services and sites, mask user access patterns, and help user devices with encryption of user data and traffic, in particular, to address the issue of insecure or malicious access routers; use of **trusted hardware** can help with identification of users (client-side) as well as with increasing trust in Secure Intermediary (via attestation of its functionality).

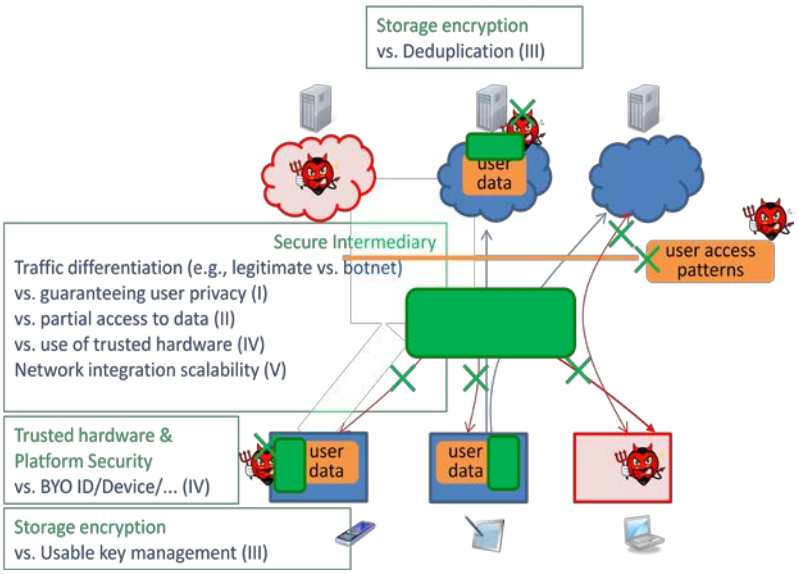

Figure 2 Challenges in addressing threats in Cloud Services

Challenges in securing Cloud Services: However, naïve application of such solutions brings in several new challenges (Figure 2) that we propose to address in this project.

I. Secure Intermediaries need to differentiate between legitimate and unauthorized traffic while still preserving user privacy. Also, it has to be ensured that the Secure Intermediary itself does not transform into a security risk, e.g. for denial of service – attacks. In particular, each Secure Intermediary must be demonstrably incapable of mining user access patterns. Merely encrypting the communication channel is not enough to prevent an adversary from extracting information about the user [17][19]. To this end, we need to use **private information retrieval** (PIR) techniques on one hand. On the other hand, related to threat (4), techniques are needed for **detecting covert/illicit channels.** These channels may be used, e.g., for unauthorized tracking of user actions or communication between nodes in a botnet. How to resolve this apparent contradiction between privacy and security requirements is an open question. Furthermore, data protection, whether content encryption or the use of trusted cloud services, adds processing and transmission delays that are not beneficial in view of the current Internet services [9]. Thus, we need to carefully evaluate what individual content and user action ultimately must be protected while taking the penalty of increased latency into account.

II. In the analysis of objects and traffic, significant parts of the relevant content and data may not be available due to technical limitations of certain closed platforms (e.g., Windows Phone or iOS), user privacy concerns, or because those parts are encrypted. **Detecting malicious or unwanted objects and actions with** a low False Positive Ratio and only partial access to the data requires sophisticated data analysis methods and adaptive logic distributed between the Secure Intermediary and endpoints.

III. To be viable, client-side encryption of user data has to be reconciled with various business needs of cloud storage providers as well as usability needs of end users [18]. For example, cloud storage providers **deduplicate** data so as to avoid storing multiple

copies of the same data objects. **Naïve** application of client-side encryption will thwart deduplication. Reconciling deduplication and client-side encryption is an active research topic [1]. Similarly, end users need to access their data from multiple devices. With strong encryption (i.e., using randomly generated keys rather than deriving them deterministically from user passwords), appropriate key management techniques need to be in place. **User-friendly key management** is an open research question. Finally, **searching** data (e.g., shared in groups) for keywords or intersections with other data sets is a related facet of importance [18].

IV. It is well understood that **trusted hardware** and trusted computing technology in general, can be used for attestation of (server) configuration and for supporting user identification. The use of *server-side* trusted hardware could lead to the possibility of simple, alternative solutions for the problems of PIR and deduplication. For example, after a client uploads an encrypted file to a server it can establish a secure communication channel to server-side trusted hardware using its certified public key and the client's own key. The client can then communicate the hash of the plaintext file and the encryption key via this channel. If server-side hardware determines that the file needs to be deduplicated, it will return the encryption key related to the already existing copy of the file. This comes with at least two challenges: First, how to *securely* integrate trusted hardware optimally with the advanced protocols we will develop? Second, how to manage user identification based on trusted hardware when users bring their own devices (BYOD) and receive their trusted identities from disparate sources (BYOI)?

V. High scalability and performance of Secure Intermediaries serving large numbers of simultaneous connections around the globe are crucial but difficult to achieve. Appropriate load balancing, communication and processing optimization techniques are required for providing sustainable and high quality security services. These must integrate well in spite of the stringent security requirements that limit the applicability of traditional load balancing and performance improvement techniques (e.g., content-based/service routing, caching). Service availability and resilience to node or path failures must also be preserved. Furthermore, intermediaries and indirection services incur changes in communication latency that must be minimized as many of the important Internet services and applications are delay sensitive [9].

State of the art. Recent research has partially addressed challenges I-V; however, a unified and encompassing solution that addresses the cloud computing environment with distributed functions as depicted in Figure 2 has not yet been developed. The key results in related work include the following:

*Deduplication and data encryption***:** Bellare *et al.* recently proposed using "convergent encryption" as a way to address the problem of reconciling client-side encryption with deduplication by deriving the encryption key from the message itself. To guard against brute-force probing attacks, they propose [1] using an auxiliary key server that imposes rate limiting. We will explore ways to avoid the use of additional servers or rate limiting.

*Computationally Private Information Retrieval (CPIR)***:** CPIR schemes based exclusively on number-theoretic assumptions are currently impractical. But recent schemes based on other assumptions hold more promise for practical deployments [2]. However, all of such single-server CPIR schemes need to process the whole database to answer a query, which increases the computation overhead to a large extent. Error-correcting codes help in ensuring reliable transmission of information over noisy channels as well as in providing reliable storage of information on a medium that may be partially corrupted. Such codes allow adding

redundancy to messages, encoding them into longer bit strings, called "codewords," in a way that the message can still be recovered even if a certain fraction of the codewords are lost or corrupted. CPIR schemes are intimately related to a special class of error-correcting codes called "locally decodable codes" (LDCs), which provide efficient random-access retrieval and high noise resilience by allowing reliable reconstruction of an arbitrary bit of the message from looking at only a small number of randomly chosen codewords. As a result, LDCs can be applied to achieve computationally efficient CPIR schemes [10][11][12], but those still require multiple servers.

Covert/Illicit channel detection: Combining analysis of illicit communications and privacy requirements has been studied by Afanasyev *et al.* [5], while Grusho *et al.* [8] have opened new angles in modeling covert channels and their detection. There is a permanent confrontation between design of covert channels and their detection; therefore new detection mechanisms are still needed.

Trusted hardware: Combining attestation functionality with end-to-end cryptographic protocols [3] is an instance of protocol composition that is susceptible to man-in-the-middle attacks [4]. We will develop trusted hardware based solutions for CPIR and deduplication that are resistant to man-in-the-middle attacks.

Data analysis: Detecting malicious or unwanted objects requires accurate, fast and adaptive classification methods. Such methods are beginning to appear [7]. The specific requirement of a low False Positive Ratio demands specific learning methods, which can also address the particular structure of the data: partial access, and new Euclidian metrics (non-Euclidian distances).

Load balancing and performance: The performance of cloud platforms is a very active research topic with many recent contributions in network/VM scheduling, data transport, and in-network processing [6]. But with new developments in cloud-based services, we have to keep the latencies to access content at a reasonable level and even seek further reduction [9] while preserving stringent security properties, which imposes additional design constraints.

In addition to the open questions identified above, building a unified system incorporating these different enablers has remained elusive.

Prior research by PIs: Aalto Comnet, Aalto CSE, University of Turku teams have previously collaborated on design and analysis of cryptographic protocols, e.g. [4], which provides some background for Challenges I and III. Aalto ICS and F-Secure have closely collaborated in D2I SHOK program on website classification for parental control purposes (Challenge II). University of Helsinki and Aalto CSE teams collaborate with TU Darmstadt as part of the Intel Collaborative Research Institute for Secure Computing (http://www.icri-sc.org) focusing on usable mobile security. We will leverage this collaboration in developing a usable key management scheme as part of challenge III. The doctoral dissertation of Dr. Jan-Erik Ekberg (Trustonic) [16], advised by Prof. N. Asokan (Aalto CSE), laid out the fundamentals of using trusted hardware for securing applications and services. We will build on the solutions from [16] to address challenge IV. Aalto Comnet and University of Helsinki have experience in designing robust and scalable systems (not just) for data centers and in characterizing performance of Internet services (Challenge V).

3 Objectives of research

In addressing the challenges I-V, our key research questions are the following:

1. What is the design space for efficient CPIR using (a) cryptographic techniques and/or (b) trusted hardware?

2. How can we design a practical cloud storage encryption service allowing (a) the user to access the data from multiple devices and (b) the storage service to perform data-driven functions such as deduplication?
3. Is it possible to design a mechanism for detecting covert/illicit channels from or to end user devices while guaranteeing privacy of legitimate user interactions?
4. How to realize adaptive security logic distributed between Secure Intermediary and endpoints? How to build a classifier that can reduce the number of false positives while keeping good global performances and optimizing a given time to performance ratio?
5. How to design a scalable high-performance networking architecture for the Secure Intermediary, building on emerging cloud technology?

The research questions will be explored in parallel developing the theory and understanding real-world deployment issues through prototype experimentation and close collaboration with the industrial partners.

Vision and Hypothesis: We envision a unified architecture for security services delivered via clouds. Our overall hypothesis is that it is feasible to solve questions 1-5 in such a way as to enable the design of a scalable, reliable, low latency system addressing challenges I-V.

We envision a future security service, the Secure Intermediary, will facilitate, together with the communication endpoints, secure and privacy-aware access of cloud and Internet resources across connection types and devices while providing coherent user experience across all popular platforms in elastic and adaptive ways.

The Secure Intermediary will help users and their devices steer away from unsafe situations in a variety of ways. The endpoints can authorize the Secure Intermediary to distinguish between legitimate and malicious/unwanted traffic objects and actions (Research Questions 3 and 4). They can also query the Secure Intermediary in a privacy-preserving manner for reputation information of URLs or applications (Research Question 1) or seek its assistance for strengthening the keys used for client-side storage encryption (Research Question 2). Emerging new technologies, such as SDN and cloud, enable real-time or near real-time traffic and content analysis, and decision-making regarding data transmission, processing and storage. The Secure Intermediary will have intrinsic scalability properties built on cloud capabilities, such as virtualized resources, while offering value-added security features for cloud services through intelligent security function placement and state management (Research Question 5).

The new and far reaching element of the research is the vision of a security control plane that has knowledge of the end points and their environment, the network and the processing elements in the cloud. The control plane together with the necessary analysis and decision-making building blocks then determines the optimal parameters in order to meet the security, privacy and performance requirements. This integration of distributed security functions, network, storage, and tasks running in virtual machines paves way for the next generation of security enhanced cloud services in the Internet and in wireless (e.g., 5G) core networks.

Relevance to the ICT 2023 call: This project will address a timely research topic: we will identify and solve critical security and privacy problems in delivering security services via clouds and in using cloud services. We aim to develop a unified solution across Internet & cloud technologies and emerging wireless systems like 5G, with the aim of performing the correct security function at the most suitable point in the network in a coordinated fashion.

International co-operation: We intend to have intensive collaborations with leading (security) research groups around the world, active at the intersection of cloud and security. We detail our international collaboration plans in Section 6.3.

Business collaboration: The CloSe project provides an important opening for the Finnish security industry to take a leadership position in this evolving area. Our consortium includes three industrial partners who represent different sectors in the Finnish security industry: a multinational corporation, Nokia Solutions and Networks (NSN), providing core network infrastructure; a medium sized company, F-Secure, offering security functions for cloud services; and the Finnish subsidiary of a small European company, Trustonic, providing Trusted Execution Environments. The industry partners will take a very active role in the research itself and the subsequent technology transfer.

Added value from consortium collaboration: The academic partners in the consortium bring complementary expertise to the project: applied cryptography (University of Turku and Aalto CSE), platform security (Aalto CSE), data analysis (Aalto ICS), and networking technologies (Aalto Comnet and University of Helsinki). Arcada University of Applied Sciences will collaborate closely with the academic and industrial partners to facilitate realization of research results. The motivating usage scenarios and the problems addressed were formulated jointly with the industrial partners based on their business needs. The teams of the academic and the industrial partners will cooperate closely in addressing challenges I-V.

4 Research methods and material, ethical issues

Research methods: Several different research methods are employed in the project, namely cryptographic protocol design and analysis, system design, and machine learning and data analysis. The theoretical designs will be realized as prototype implementations in a laboratory environment to study their properties.

Research material consists mainly of data collected by our industrial partners. Data will be collected based on the work package constraints from each industrial partner under the guidance of Arcada School of Applied Sciences. Data will be stored at each industrial partner site, using secure access with strong cryptographic protection; access will only available to the PIs of the project. In addition, the data (and any accompanying metadata) will be encrypted using personal PGP keys generated for each PI of the project. Therefore, only PIs will have the original access to the data sites. Handling of the data at each site will then be under the responsibility of the local PI, taking appropriate measures to pass and store the data to the researchers under his supervision.

Ethical considerations: All the datasets collected by the three industrial partners (assisted by Aalto University and Arcada School of Applied Sciences) will be automatically anonymized and pre-processed at the collection site, in order to fully guarantee the personal rights of the customers/clients/collaborators (denominated as "Test Group") of the above-mentioned partners. Data will be collected from a sample of the selected Test Group with their explicit *a priori* consent. Gathered data will be stored securely. We will ensure limited, but sufficient, access for project partners. In any case, this arrangement will not impede the publication of the scientific results in conferences and/or journals.

Risk management: Throughout the project a steering group consisting of the PIs, with the help of the scientific advisory board will monitor project progress and correct course as needed. Detailed information about project governance is provided in Section 5.4. A detailed risk management plan is presented in Section 8.2.

5 Implementation: Timetable, budget and distribution of work

We expect the project to proceed in three phases as shown below over the 2-year duration:

	Duration	Description	Milestone
Phase 1	Y1M1-Y1M3	Matching state of the art and real world requirements (with industry partners)	MS1 (Y1M3)

Phase 2	Y1M2-Y2M8	Designing, implementing and evaluating solutions for Research questions 1-5	MS2 (Y1M12), MS3 (Y2M6)
Phase 3	Y2M1-Y2M12	Designing a unified system architecture addressing Challenges I-V (addressing the research hypothesis)	MS4 (Y2M12)

Results will be published at each of the four milestones at a results-sharing workshop that coincides with the milestone.

5.1 Work Packages

The project will be structured into five thematic work packages, each work package corresponding to a research question identified in Section 3. In addition, there is a unifying work package. Each work package consists of one or more tasks.

5.1.1 Work Package 1: Computationally Private Information Retrieval (CPIR)

This work package addresses the question of CPIR: how to allow a user client to interact with a server database without revealing information about the queries or responses to the server or any third party. Our goal, unlike previous work, is to design a single-server CPIR scheme that is efficient both in communication and computation, without offline computation. One possible approach here that we intend to study the combination homomorphic encryption schemes (possibly, based on coding theory or lattices) to locally-decodable codes. However, since our main goal is efficiency, we will study several competing approaches.

The tasks and deliverables of this work package are as follows:

Task	Name	Deliverables
1.1	Requirements analysis for PIR	Requirements document (MS1)
1.2	Efficient single-server PIR scheme	Design (MS2), prototype (MS3)
1.3	PIR using trusted h/w	Design (MS3), prototype (MS4)

Participants: Aalto CSE, University of Turku
Partners: F-Secure, Trustonic (industrial); University of Tartu (international)

5.1.2 Work Package 2: Secure cloud storage and deduplication

This work package addresses the question of designing and implementing a practical, strong client-side encryption for cloud storage that allows (a) the storage provider to perform deduplication when necessary and (b) users to access stored data from multiple client devices. For "(a)", our objective is to use a standard semantically secure (but not convergent) encryption scheme for file encryption but design a non-interactive, privacy-preserving key management protocol that allows uploaders of the same file to share the same encryption key. In contrast to previous work, our goal is to design the key management protocol without using any additional servers. For "(b)", our objective is to design an easy-to-use key management interaction that allows a user to access her files from all her personal devices. This would involve designing a key hierarchy for secure storage and using out-of-band channels (like NFC, or QR codes and cameras) for key management.

The tasks and deliverables of this work package are as follows:

Task	Name	Deliverables
2.1	Requirements analysis for secure cloud storage	Requirements document (MS1)
2.2	Key management for secure cloud storage	Privacy-preserving key management protocol between different users: design (MS2), prototype (MS3) User-friendly key management mechanism for personal devices of a user: design (MS2), prototype (MS3)
2.3	Deduplication with trusted h/w	Design (MS3), prototype (MS 4)

Participants: Aalto Comnet, Aalto CSE
Partners: F-Secure, Trustonic (industrial); TU Darmstadt, Univ. of Tartu (international)

5.1.3 Work Package 3: Detecting covert/illicit channels

This work package addresses the issue of detecting covert/illicit channels from or to end user devices while at the same time guaranteeing privacy of legitimate user interactions. The Secure Intermediary cannot itself be able to find out user access patterns but it should still be able to distinguish between legitimate user traffic from unauthorized traffic (covert channels from malicious apps on user devices or encrypted Command & Control traffic for botnets). New mechanisms are developed to achieve these somewhat conflicting goals, based on modeling of covert channels and their detection.

The tasks and deliverables of this work package are as follows:

Task	Name	Deliverables
3.1	State of the art and requirements	Requirements document (MS1)
3.2	Privacy-preserving detection of covert/illicit channels	Design document (MS2), prototype (MS3)
3.3	Integrating detection mechanisms as part of Secure Intermediary	Design (MS3), prototype (MS4)

Participants: University of Turku
Partners: F-Secure, (industrial), Xidian University (international)

5.1.4 Work Package 4: Traffic differentiation

This work package focuses on the analysis of objects and traffic, for the detection of malicious or unwanted objects with low False Positive Ratio. The main challenges are the partial availability of the raw data, either for technical limitations of the running platforms (e.g., mobile phones) as well as the possibility of the data to be encrypted or obfuscated. This requires sophisticated data analysis methods and adaptive logic distributed between Secure Intermediary and endpoints. Specifically, the data extraction and analysis must be able to run on both the backend (server-side) as well as on the end user device. The capabilities of the hardware being very different, the method must have several levels of computing requirements as well as results expectations.

The tasks and deliverables of this work package are as follows:

Task	Name	Deliverables
4.1	Data collection and full-scale data analysis,	Document on available data, with analysis on usability of data for processing (MS1)
4.2	Low requirements model/app for mobile devices	Document/Results report (MS2), prototype apps and models (MS3)
4.3	Adaptive capabilities of the models/apps	Final software for backend (MS3), final apps for devices (MS3), Final results document (MS3)

Participants: Aalto ICS, ARCADA
Partners: F-Secure, NSN (industrial), University of Iowa (international)

5.1.5 Work Package 5: Networking and scalability

This work package addresses the scalability and performance requirements for the Secure Intermediary service that connects the endpoints with the cloud environment. We will design and evaluate a distributed Secure Intermediary that is able to flexibly interact and integrate with cloud services. It will allow security logic and state to be instantiated and maintained at the appropriate locations along the end-to-end path between the user and the service. We will devise appropriate techniques for achieving elasticity and scalability, adapting means for load balancing, VM mobility, service chaining, and caching to match our specific constraints. The work package will explore using current trends in scalable cloud network (e.g., Software-defined Networking, SDN) and cloud platform technologies (e.g., OpenStack) to implement efficient, scalable, robust, and cost-effective solutions.

The tasks and deliverables of this work package are as follows:

Task	Name	Deliverables
5.1	State of the art survey and requirements definition	Tools and gap analysis and requirements specification (MS1)
5.2	Design and implementation of a distributed Secure Intermediary and integration with cloud infrastructure	Design specification and prototype implementation (MS2) Integration of Secure Intermediary with cloud infrastructure (MS3)
5.3	Evaluation of Secure Intermediary	Experimentation with Secure Intermediary: performance (MS 3) and scalability (MS4)

Participants: Aalto Comnet, University of Helsinki
Partners: F-Secure, NSN (industrial), University of Cambridge (international)

5.1.6 Work package 6: Unified design

Based on the insights gained from all previous work packages, this work package will generate a proposal for a unified system architecture.

Task	Name	Deliverables
6.1	Unified system architecture	Design document (MS4)

Participants: all academic and industrial partners.

5.2 Sub Projects

The participants in the projects are organized into sub-projects, one per partner. The table below lists the names of the researchers expected to be contributing to the project and the extent of their contributions to the work packages. Some are already employed by the participating institutions, and the rest will be recruited when the project begins. We expect to employ a total of **9 full-time researchers**, each contributing 2 person years during the project period. The PIs of each sub-project will devote 3 person-months in leading the project.

Sub-project	Researcher	WPs
CloSe-Aalto-Comnet		
	Jörg Ott (PI)	2,5,6
	Arseny Kurnikov (PhD student)	5,6
	NN (hiring in progress) (postdoc)	2,6
CloSe-Aalto-CSE		
	N. Asokan (PI)	1,2,6
	Jian Liu (PhD student)	1,2,6
	Sandeep Tamrakar (PhD student)	2,6
CloSe-Aalto-ICS		
	Juha Karhunen (PI)	4,6
	Yoan Miche (co-PI)	4,6
	Luiza Sayfullina (PhD student)	4,6
CloSe-ARCADA		
	Amaury Lendasse (PI)	4,6
	Kaj-Mikael Björk (co-PI)	4,6
	Emil Eirola (postdoc)	4,6
CloSe-UH		
	Sasu Tarkoma (PI)	5,6
	Aaron Yi Ding (postdoc)	5,6
CloSe-UTurku		
	Valtteri Niemi (PI)	1,3,6
	Tommi Meskanen (researcher)	1,3,6
	Noora Nieminen (PhD student)	1,3,6

5.3 Project meetings and governance

Each work package will meet frequently for technical discussions. Each milestone will be accompanied by a workshop where results will be presented publicly. A steering group consisting of the PIs and representatives of the three industrial partners will meet during the workshops (and as needed otherwise) to evaluate project progress. At the beginning of the project the steering group will invite selected world-class experts to serve in the scientific advisory board (SAB) for the project. The SAB will be invited to a results workshop once a year for feedback and guidance.

6 Researchers, research environment, mobility

6.1 Research teams of sub-projects

CloSe-Aalto-Comnet: Aalto University, Department of Communications and Networking team is led by Prof. Jörg Ott (http://www.netlab.tkk.fi/u/jo/) and includes one post-doctoral researcher and one PhD student. It has expertise in protocols, paradigms, and systems for distributed content storage and retrieval techniques, in security properties of content-centric networks, and in secure interactions between peers without revealing private information. A recent focus is on censorship-resistant communication. Prof. Ott is member of ACM, IEEE, and GI, treasurer of ACM SIGCOMM, vice chair of IEEE Comsoc TCCC, and on the editorial board of Elsevier Computer Communications.

CloSE-Aalto-CSE: Aalto University, Department of Computer Science and Engineering team (http://cse.aalto.fi/en/research/secure-systems/) is led by Prof. N. Asokan and includes a post-doctoral researcher and a PhD student. It brings expertise in applied cryptographic protocols, platform security 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 ACM TISSEC, the premier security research journal. He was recently granted a Google Research Award.

CloSE-Aalto-ICS: Aalto University, Department of Information and Computer Science team EIML (Environmental and Industrial Machine Learning) (http://research.ics.aalto.fi/eiml/) is led by Prof. Juha Karhunen and includes a PhD student. It brings expertise in Data Analysis and Machine Learning for industrial applications. Prof. Karhunen is a world class expert in Independent Component Analysis and his book is the reference in the topic (more than 10500 citations). He is leading another research group on Deep Learning and Bayesian Modeling (see http://research.ics.aalto.fi/bayes/). Deep learning has provided world record results in many benchmark machine learning problems and is currently a hot topic in machine learning.

CloSe-ARCADA: Arcada University of Applied Sciences (http://www.arcada.fi/en) team is led by Principal Lecturer Amaury Lendasse and includes a post-doctoral researcher. Dr. Lendasse is also a Docent at Aalto University, ICS department and a former Ikerbasque Professor at the University of the Basque County. He is a world class researcher in Time Series Prediction and Industrial Machine Learning. He was the PI in several industrial projects for a total funding of 2 MEUR.

CloSe-UH: University of Helsinki team (http://www.cs.helsinki.fi/en/nodes) is led by Prof. Sasu Tarkoma and includes a post-doctoral researcher. The team brings expertise in cloud platforms, networking, and security of communication protocols. Prof. Tarkoma has written three books on mobile and Internet technology and he has published over 120 scientific articles. He is a senior member of IEEE and editorial board member of Elsevier Computer Networks journal.

CloSe-UTurku: University of Turku team (http://tucs.fi/research/research-units/fundim/) is led by Prof. Valtteri Niemi and includes a post-doctoral researcher and a PhD student. The team has expertise in cryptography, its applications and security of mobile communications.

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.

Research infrastructure support from sub-project sites: The project will make extensive use of research and computing infrastructure in the participating universities. For example, University of Helsinki has a 2000-core data center (Ukko) and a production quality OpenStack deployment that will be utilized in the project. All the resources combined, we have experimental facilities for end-to-end experiments that include the mobile access network, mobile core network, datacenter, and cloud platform.

6.2 Industrial Partners

F-Secure's objective in the project is to develop novel ways of protecting user privacy, data, and devices via cloud-based security services. With the increasing use of mobile devices to access the Internet for both private and business purposes, people are facing major threats, such as malicious objects and content, on-line tracking, compromised public Wi-Fi networks, and so on. Since purely on-device protection mechanisms are inadequate for solving such security problems, F-Secure plans to design, implement, and validate methods based on the "secure intermediary in the cloud" paradigm. From their point of view, important project objectives are: providing coherent user experience across all popular platforms; elasticity and adaptability of the services; addressing the needs of both individuals and businesses.
NSN 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 monitoring system that utilizes artificial intelligence for recognizing malicious traffic and usage patterns and that scales 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 building of the final combined architecture proposed in the project.

6.3 International Academic Partners

Partner	Description	WPs
TU Darmstadt (Prof. Ahmad-Reza Sadeghi)	User-friendly key management for personal devices of a user.	2
University of Tartu, Institute for Computer Science (Dr. Helger Lipmaa)	CPIR; Privacy-preserving key management protocol between different users.	1, 2
University of Iowa (Prof. Nick Street)	Embedded systems and Big Data	4
University of Cambridge (Prof. Jon Crowcroft)	SDN for security enforcement	5

The project has an excellent research network that includes both national and international research groups. The **international** key collaborations are listed in the table. There are also active collaborations with UC Berkeley and ICSI (US), UC Irvine, Nanyang Technological University (Singapore), XiDian University and Tsinghua University (PRC). The key **national** collaborator is Prof. Kaisa Nyberg, who is part of a consortium for privacy-aware retrieval and modelling of genomic data (PRIGENDA). They develop privacy-preserving machine learning that can make use of some of the same techniques, like CPIR, that we need to use. However, their intended applications areas differ from ours. Therefore their requirements (in terms of required level of confidentiality and performance) differ and hence will admit different solutions. We will liaise closely with Prof. Nyberg and PRIGENDA.

6.4 Other Collaborations

DIGILE: Via collaboration with DIGILE, CloSe research will contribute to better understanding of the key security challenges related to the wide and active adoption of cloud services and will also help identify technology and business opportunities for Finnish industry to play a significant role in that global process.

CSC: University partners have access to Grid Computing facilities in the context of the NorduGrid and the Enabling Grids for E-sciencE (EGEE) initiatives. The EGEE infrastructure in Finland is managed by CSC, the Finnish IT centre for Science. Aalto University currently uses the CSC grid facilities and has access accounts with them. The CSC grid structure (ranked 336 in the TOP500 ranking of SuperComputers) is composed mainly of two instances: *Sisu* is a Cray XC30 supercomputer consisting of 1472 2.6 GHz E5-2670 CPUs, each containing 8 cores with an eventual computaional capacity expected to surpass petaflops per second; *Taito* is a HP Proliant SL230s Supercluster consisting of 576 HP ProLiant SL230s servers each equipped with two Intel Xeon 2.6 GHz E5-2670 CPUs totaling in 1152 processors or 9216 cores.

6.5 Researcher Mobility

We intend to have two-way mobility between industrial partners and academic research groups in the project: we plan for multiple summer 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. Moreover, we have agreed on several research visits by project researchers to groups of international collaborators as well as to host visits by collaborators to Finnish institutions.

Outgoing mobility:

Partner	Participant	Sub-Project	Description
TU Darmstadt	Sandeep Tamrakar	CloSe-Aalto-CSE	Two week visit (2015), WP2: user-friendly key management
U. of Tartu	Jian Liu	CloSe-Aalto-CSE	2 x one month visits (2014, 2015), WP1,WP2: cryptographic schemes for CPIR and deduplication.
Cambridge U.	Aaron Yi Ding	CloSe-UH	2 x one month visits (2014, 2015), WP5: Collaboration on SDN for security enhancement**:**
U. of Iowa	Luiza Sayfullina	Close-Aalto-ICS	One month visit (2015), WP4: Computational time improvements for backend classifier training (on Big Data)
U. of Iowa	Emil Eirola	CloSe-ARCADA	Two months visit (2015), WP4: Establishing relationship between confidence levels on decision, model parameters and user settings.

Incoming mobility:

Partner	Participant	Sub-Project	Description
U. of Tartu	Helger Lipmaa	CloSe-Aalto-CSE	One month visit (2015), WP1,WP2: cryptographic schemes for PIR and deduplication.
Xidian University	Mingjun Wang	CloSe-UTurku	Two months visit (2015) WP3: privacy vs covert channel detection.

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. Also, they plan to have several summer intern or Master thesis worker positions for the research partner groups during the project.

7 Researcher training and research career

The core research team consists of 4 PhD students and 5 post-doctoral researchers. The PhD students will do their dissertation research in this project. 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. The PhD students are expected to participate in doctoral schools – e.g., the Helsinki ICT Doctoral Education Network (HICT), University of Turku Graduate School or the EIT ICT Labs Doctoral School – to broaden their expertise and offer forums for interactions.

Two of the nine researchers expected to be hired for the project (Nieminen and Sayfullina) are women. All academic and industrial partners are committed to support and value diversity and equal opportunity of individuals working in/with their respective organizations.

8 Expected research results, possible risks

8.1 Scientific and societal impact

The expected results include high-quality scientific articles addressing the challenges and research questions, prototype implementations of the proposed techniques, a unified design 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-based Internet services for which security is a crucial requirement.
Concrete research results for the research question 1 are expected to include novel methods that use each enabler, combinations of these methods, comparative analysis and implementations. Solutions to questions 2 and 3 are expected to be concepts, mechanisms, possible trade-offs between different goals, implementations and performance analysis. Similarly, solutions to question 4 are likely to include novel concepts, methods, optimizations, and implementations. The design to address question 5 is expected to provide mechanisms for scalable distributed security functions. Finally, the unifying thesis is expected to be solved by system architecture and related concepts, various mechanisms, prototype implementations and thorough analysis.

8.2 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	Likelihood	Mitigation
1	No efficient single-server CPIR is found	1.2	Medium	a) Increase emphasis on CPIR using trusted hardware (Task 1.3) b) Prototype a known scheme [2]
2	Crypto protocol deduplication key management does not meet all requirements	2.2	Low	Increase emphasis on key mgmt using trusted hardware (Task 2.3)
3	Not possible to find efficient trade-off between privacy requirements and covert channel detection	3.2	Medium	Decrease privacy level in controlled manner for nodes/traffic where probability of botnet presence is estimated to be high.
4	Encrypted data cannot be used as such	4.2	Medium	Have some decryption (traffic, metadata, data) done safely (i.e. on a secure backend).
5	Low feature set turns out to be not good enough, e.g., for acceptable FPR	4.2	Low	Find larger data sets for obtaining adequate performance (may require more CPU power and bandwidth and faster feature extraction methods, e.g. extreme learning machine - [14][15] and deep learning [13].)
6	The full set of Secure Intermediary functionality is known only near to the end of the project.	5.3	Low	Consider network integration regularly throughout the project

8.3 Applicability of research results

The three industrial partners (F-Secure, NSN, 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). The industrial partners plan to have their research teams work closely together, in particular, by inviting team members of the academic partners to work in the company environments and by establishing internship and thesis worker positions. They believe that the proposed project provides excellent opportunities for conducting high-class research targeted at solving real-life problems in the domains of user privacy and data, device, and network security.

8.4 Dissemination of results

The results will be published in top conferences and journals, i.e. in fora like ACM CCS, Usenix Security, ACM SIGCOMM, ACM CoNEXT, ACM ICN, IEEE/ACM Transactions on Networking, Journal of Machine Learning Research, 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. We will organize a special session (or workshop) on Machine Learning for Security at one of the top international conference on Machine Learning. Based on this special session, a special issue in a top journal will be published.

Project results will be incorporated in the courses taught at academic partner institutions: e.g., a graduate course on Machine Learning for Security will be organized and security protocol mechanisms will be incorporated in Protocol Design at Aalto University. A joint course on Mobile System Security by Aalto University and University of Helsinki is planned to commence in Spring 2015.

Industrial partners expect that the project results will contribute to their products and services for the public. Whenever appropriate, they are planning to communicate that such services utilize approaches and technologies developed via the CloSe research work.

9 Key literature or bibliography

[1] M. Bellare, S. Keelveedhi, and T. Ristenpart: DupLESS: Server aided encryption for deduplicated storage, Usenix Security Conference, 2013.

[2] F. Olumofin, I. Goldberg: Revisiting the Computational Practicality of Private Information Retrieval, Financial Cryptography and Data Security, 2012.

[3] Y. Gasmi, A.-R. Sadeghi, P. Stewin, M. Unger, N. Asokan: Beyond secure channels. Scalable Trusted Computing (STC), 2007.

[4] N. Asokan, V. Niemi, K. Nyberg: Man-in-the-Middle in Tunnelled Authentication Protocols, Security Protocols, 2003.

[5] M. Afanasyev, T. Kohno, J. Ma, N. Murphy, S. Savage, A. C. Snoeren, G. M. Voelker: Privacy-preserving network forensics. Communications of the ACM 54(5): 78-87 (2011)

[6] J. Sherry et al: Making middleboxes someone else's problem: network processing as a cloud service. Proc. ACM SIGCOMM 2012: 13-24

[7] Y. Miche, A. Akusok, J. Hegedüs, R. Nian and A. Lendasse: A Two-Stage Methodology using K-NN and False Positive Minimizing ELM for Nominal Data Classification, In Cognitive Computation. 2014, published online, DOI: 10.1007/s12559-014-9253-4

[8] A. Grusho, N. Grusho, E. Timonina: Problems of Modeling in the Analysis of Covert Channels. Proc. Int. Conf. on Mathematical methods, models and architectures for computer network security, 2010. http://dl.acm.org/citation.cfm?id=1885205

[9] Internet Society, Workshop on Reducing Internet Latency, 2013 http://www.internetsociety.org/latency2013

[10] S. Yekhanin: Towards 3-query locally decodable codes of subexponential length, Journal of the ACM (JACM), 55(1): 1 (2008).

[11] K. Efremenko: 3-query locally decodable codes of subexponential length, SIAM Journal on Computing, 41(6): 1694-1703 (2012)

[12] P. Raghavendra: A note on Yekhanin's locally decodable codes, Electronic Colloquium on Computational Complexity (ECCC), TR07-016. 2007

[13] K. Cho, T. Raiko, A. Ilin, J. Karhunen: A Two-Stage Pretraining Algorithm for Deep Boltzmann Machines. ICANN 2013, Lecture Notes in Computer Science, vol. 8131, pp. 106-113, Springer, 2013. http://users.ics.aalto.fi/juha/papers/icann13jp.pdf

[14] Y. Miche, A. Sorjamaa, P. Bas, O. Simula, C. Jutten, A. Lendasse., OP-ELM: Optimally-Pruned Extreme Learning Machine. In IEEE Transactions on Neural Networks, volume 21, pages 158--162. January, 2010. http://dx.doi.org/10.1109/TNN.2009.2036259

[15] E. Cambria, *et al.*: Extreme Learning Machines [Trends Controversies]. In Intelligent Systems, IEEE, 28(6), Pages 30-59, 2013.

[16] J-E. Ekberg: Securing Software Architectures for Trusted Processor Environments, D.Sc dissertation, Department of Computer Science and Engineering. May 2013. http://lib.tkk.fi/Diss/2013/isbn9789526036328/isbn9789526036328.pdf

[17] B. Miller, L. Huang, A. D. Joseph, J. D. Tygar: I Know Why You Went to the Clinic: Risks and Realization of HTTPS Traffic Analysis, 2014. http://arxiv.org/abs/1403.0297

[18] R. A. Popa et al: Building web applications on top of encrypted data using Mylar, to appear in NSDI'14 (11th USENIX Symposium on Networked Systems Design and Implementation), 2014.

[19] Ning Xia et al: Mosaic: Quantifying Privacy Leakage in Mobile Networks. Proc. ACM SIGCOMM conference, August 2013.