Security and Privacy:

From the Perspective of Software, Services, Cloud and Data

NESSI White Paper

Networked European Software and Services Initiative

March 2016

Executive Summary

NESSI is the Horizon 2020 European Technology Platform (ETP) for software, services and data. This white paper focuses on the increasing role of security and privacy and highlights research directions from a NESSI perspective. The paper does not aim to look at security and privacy from all angles, but on the impact on software, services, cloud and data of all kinds.

The paper starts with an introduction focusing on key emerging challenges and focuses on business, application sectors, education and law, regulation and policies. Research opportunities are the core of this white paper. The main findings related to the areas covered in this document are listed below. Additional research items and more detailed information are provided in the relevant sections.

- **Managing risks** – Tools to improve the analysis and treatment of risks by reducing the reliance on labour intensive and potentially error-prone analysis by experts is required due to e.g. regulatory compliance requirements, data exchanged across organisations. Standardization of protocols and security intelligence data formats to facilitate interoperability would enable more reliable methods to analyse risks related to security and privacy, also enabling automated assessment and comparison.
- **In the cloud** – Privacy-enhancing technologies (PETs) and cryptographic mechanisms are to be enabled for applications in cloud environments. Standardized frameworks would support making transparency-related information available to cloud customers and would enable automated security and privacy enforcement and traceability in federated clouds. In addition research is required to better communicate information, help users understand the privacy consequences of their online actions, and support users in taking decisions.
- **While analysing data** – The risk of re-identification of persons rises with the number of information items. Therefore research on anonymization techniques and further insights about linking parameters is required. Technologies exist to detect security incidents and violations; little exists in the area of privacy. Statistical privacy, research on implementing measures directly to the core of the system and a categorization and formulation of principles is also needed.
- **Awareness in life cycle** – The environment in which a service operates may change. Therefore, the software should also be monitored at runtime for any security or privacy breaches. Only a few security engineering practices developed specifically for agile processes exist. Privacy-enabled applications must be ‘certified’ or ‘labelled’ to existing regulations and legal requirements. Clearly defined automatic policy handling could be achieved, if data are equipped with sticky policies.
- **Data analytics for security and privacy** – Research should be focused on the feasibility of using analytics software as a kind of ‘early warning system’ to detect whether data could be misused. Analytics could automatically detect, identify and classify sensitive information. Existing privacy technologies should be improved such that they become useful and usable by guaranteeing performance, scalability, availability, and utility.

Finally the white paper highlights some example application sectors in which NESSI Partners are active.

Contents

1. Introduction

1.1. Focus of the White Paper

NESSI aims to have an impact on the technological future and innovation capabilities by identifying strategic research directions and proposing corresponding actions in our main areas such as Software Engineering, Software-based Services, Big Data and Cloud. NESSI is a European Technology Platform (ETP) active at an international level (see http://www.nessi-europe.eu/).

This white paper identifies the impact of Security and Privacy from the perspective of the main areas of NESSI and vice versa. This paper thus is not aiming to look at Security and Privacy from all angles, but at providing a complementary contribution from the NESSI point of view. In particular the paper assesses the impact on Software, Services, Cloud and Data.

1.2. Key Emerging Challenges

Security deals with the integrity, confidentiality, availability and protection of the system including data against any harmful disruption, while *privacy* deals with the protection of personal data against specific threats that harm the data subject by undermining its human right to privacy and self-determination as defined in the European Convention on Human Rights (ECHR) [1].

The main challenges of security and privacy stem from the proliferation in the use of data across distributed information systems. The sources of this data have greatly increased, along with the ways the data is being used (even to automate decisions with real-world consequences). Part of this growth is in the amount of personal data being shared (and used) in the Internet.

The main security challenge has moved from simply preventing attacks to identifying and preventing ways in which the sharing and use of all this data may cause harm. This includes the ways in which the use (or abuse) of personal data may breach privacy. New risks are emerging due to the scale of data sharing and use, e.g. the risk of unintentional harm through the use of new technologies mining large amounts of public data. This can allow de-anonymization and even prediction of personal behaviour with a high degree of probability, posing risks to privacy. If the results are used to automate other processes, this can also lead to other harmful consequences for systems, humans and society.

The increased scale and complexity of data sharing and use leads to four main challenges:

- the challenge of identifying potential security risks in complex, extended and dynamically evolving systems;
- the challenge of assessing risk impacts and the knock-on effects in such systems;
- the challenge of addressing risks needing security measures on a similar scale, e.g. risks from large scale aggregation and analysis of individually non-sensitive data; and
- the challenge for users (including data subjects, and also system designers and operators) to understand and manage security risks and protect personal data.

Security measures that cannot be understood and managed will not be effective in addressing risks. This requires that all security and privacy protection measures offer a high degree of user acceptance, transparency, and user-friendliness in their design and implementation.

1.3. Business, Application Sectors, Education, Law & Regulation & Policy

Security and privacy solutions are always integrated solutions which have to deal with policies, people, procedures and technology. Such solutions must be compatible with the business and user contexts in which they operate. Examples are the business risk and trust models associated with the use of services across organisational and geographical boundaries, the security awareness and expertise of the humans creating and using services and applications, the legal constraints under which they operate, the acceptability to individual users and user communities, and the economic cost-effectiveness of solutions used.

Software and service technology providers should consider how their offerings might incorporate the means for operators and users of the technology to maintain security and privacy at the desired level. These aspects start with the expertise to identify, understand and address risks, continue with how to embed this expertise into practice, followed by economical sustainability and acceptance considerations, and conclude with integration of the technologies.

When it comes to security and privacy there is always a trade-off between sensitiveness and functionality/usability. As an example, security and privacy by design are seen as a business advantage, rather than limiting the opportunities in the market. It is an approach to protecting privacy and security by embedding it into the design specifications of technologies, business practices, and physical infrastructures. Privacy by design (PbD) is currently included in the revised draft regulation for data protection in the EU, referred to as “Data Protection by Design” [2] in order to increase incentives to implement PbD for both suppliers of systems that process personal data and for government organisations that procure such systems.

Besides the clear technological and usability benefits of the above mentioned approaches there are also economic benefits. Studies have considered the relative cost to rectify security and privacy defects to be 1% at design time, 6.5% at implementation, 15% at testing and 100% at maintenance [3]. Therefore it is important to take security and privacy as a key enabler for the development and maintenance of trust in the digital world. Secure and privacy respecting technologies and platforms are a basis to build on. The key is to develop privacy and secure business-friendly implementations.

The implementations span industrial sectors (e.g. transport/logistics, energy, finance). The focus should be on those sectors that represent critical infrastructures. Among the most important sectors are:

- health and welfare;
- utilities and smart metering;
- intelligent transportation and logistics;
- banking and finance;
- public administration;
- defence; and
- telecommunications (not limited to 3G/4G-LTE/5G) and their protocols.

Beyond industrial or business sectors, there are technology trends that are already present in daily life and systems, and ought to be the focus of security research as well:

- Cloud Computing;
- Big Data and Data Analytics;
- Network Function Virtualisation including Software Defined Networking; and
- Cyber Physical Systems and Internet of Things.

Users and organisations are following these technology trends, and one of the most important obstacles is to overcome the lack of awareness and education of users, and even of organisations, in the field of security and privacy. In an ideal situation, the user need not take

notice of security and privacy measures, which should be completely transparent. But this means end-users do not think about security and privacy risks, and usually accept all necessary relaxation of security demanded for application use. Furthermore, in a connected world, the competition for services is global and end-users may sometime sacrifice privacy or security for good and efficient services. Nowadays data analytics, big data and cloud computing gets more and more awareness and allows constant tracking, thus it becomes more difficult to change the way technologies can encroach on our lives. To protect our privacy, we need better understanding of the purpose of the technologies and why they are valuable.

Education of users and organisations is one of the effective measures to protect personal data online, such that they will make informed choices and conscious use of Internet services when sharing their data, thus being empowered data subjects [4]. However, this doesn't mean that other measures should not be employed, including better methods for assessing risks, more user-friendly and understandable security mechanisms, and improved regulations. Users wish to use online services, irrespective of the country base of the service provider. Regulation should be adapted to the reality of Internet use. In the view of privacy, a quite general regulation is defined by the European Convention on Human Rights [5], and a more specific one by the EU Data Protection Directive [6]. The latter one is based on principles including:

- consent (of the user for processing his data);
- proportionality (only those data shall be collected which are necessary for processing and for the purpose they are claimed for); and
- marketing opt-out (the user can refuse the use of his data for marketing purposes at any time)

Although this defines a good framework for the goals of privacy protection there are several drawbacks:

- some people see this as an inhibitor for innovation;
- it might be difficult to determine whether a business process or application is compatible with the regulation;
- the implementation of compatible business processes could be too costly in the start-up phase;
- the lack of harmonisation of rules and legal norms for personal data protection among countries; and
- connection / harmonization between policy makers and ICT practitioners is required.

Due to the trend that business application providers may run their software and even their applications in any country of the world, i.e., a service may be accessed in one country but is actually executed and data stored in another country or even on another continent. Hence, there is an urgent need to develop a common vision on privacy and data protection that will push for the development of international standards, regulation and policies unifying the technical principles that guide the development of privacy-friendly systems and services.

2. Research Opportunities in the Field of Security and Privacy

2.1. Managing Risks

Systems consist of a heterogeneous mix of distributed technologies and are based on different paradigms, in order to support and enable digital business processes, and provide benefits for the economy and society. Advanced networking, cloud and Internet of Things (IoT) are examples of technologies that are ubiquitous in everyday life (healthcare, communication, retail, etc.).

Putting in place appropriate security and privacy measures requires understanding of the context and requirements (contractual, legal, etc.), identification of risks that may compromise the security and privacy properties of the system, assessment and prioritisation of these risks and their knock-on effects, and deploying effective controls and monitoring for the diagnosis of these risks.

- Currently the identification of threats is a manual process. There is need for tools to improve the analysis and treatment of risks by reducing the reliance on labour intensive and potentially error prone analysis by experts. This is especially challenging where risks may arise due to interdependency between different physical, application, business or administrative domains. This is increasingly the case as the scale and complexity of data sharing and analysis itself increases.
- An overall approach to risk management is well established based on international standards like ISO 31000, ISO 27001 and ISO 27005. However, standard approaches do not adequately handle interdependencies within and between ICT and non-ICT systems. This gap in methodology has been recognised and highlighted by the EC with respect to critical infrastructure risk management in 2012, but the problem also exists when decisions which are taken by an ICT system may have consequences beyond that ICT system. Any failure in security or privacy may then be amplified through the cascading of threat to other systems as a consequence of the initial failure. This makes it difficult to assign responsibilities, design risk mitigation strategies, or recognise which threats are the root causes of problems when they arise.
- More reliable methods are needed to analyse risks arising in interdependent ICT systems including big data systems. Methods are also needed which can aid in monitoring and diagnosis of threats to security and privacy during the system operation. Defining metrics is important to detect whether known risks have occurred however, more analysis and correlations of system and users behaviours is important to detect other zero-days threats.
- More effective analysis can be done if security intelligence data was available across organisations. Research is needed on optimal ways to express this knowledge as well as to communicate and process it effectively. Machine understandable representations may be needed to allow automated analysis and precise communication between stakeholders, e.g., through some trusted third party mechanism. This may also allow the automated assessment and comparison of the security and privacy level of different systems. This is a first step towards verifying that the composition of these systems is also secure and preserves the subsystems privacy guarantees.
- Security assurance and certification is yet another challenge. ISO27001 allows the certification of systems compliance as a whole, whereas other standards and initiatives focus on providing assurance in earlier phases of the lifecycle. The Common Criteria for Information Technology Security Evaluation (Common Criteria or CC for short), validates a “target of evaluation” against security claims made by the developer about the security properties of the product. However privacy aspects are still missing and auditing methods to demonstrate and provide evidence of their fulfilment are essential.

They should be designed and integrated in a transparent and comprehensible manner. Towards this, the development of privacy impact assessment guidelines should be further developed and integrated in the business concept, as an integral part of a privacy by design methodology. The advancement of the CSA Privacy Level Agreement (PLA) and the exploitation of ISO 29134 could be the basis for further research in this area.

- With the advancement of digitalization the number of sources of security and privacy related information increases rapidly. To keep up with the increasing amount of information new ways of collecting, storing, and processing security and privacy related data are required. Standardization of protocols and data formats is needed to facilitate interoperability and to increase economic feasibility.
- Another challenge is posed by the continuous evolution of security and privacy threats. New threats are invented with increasing speed. Hence the methods of correlation and analysis must be developed to keep the accuracy at acceptable level and to provide meaningful and useful information to human operators. A related research challenge is the development of metrics and measurements to determine similarity of data sets to allow comparison of different systems or performance of a specific system over time.
- Often it would be beneficial to extend the scope of situational awareness beyond one organisation. This is challenging as usually the security and privacy related information is considered sensitive by organizations. Therefore anonymous ways of combining the information is required. This would allow a larger view on situational awareness without revealing anything about the individual organizations.
- The advancement of information technology has introduced regulations governing security and privacy aspects of ICT systems. To demonstrate compliance with regulation is an expensive task. Therefore tools to automate gathering of evidence to demonstrate the fulfilment of relevant regulation is needed. This is important as it may affect the public perception of the service.
- Traditional Security Information and Event Management (SIEM) focuses on detecting security issues, but pays little attention to privacy. For SIEM to be fully generalised to security and privacy, monitoring mechanisms must be in place to keep track of what kind of data that is accumulated and provide means for understanding when and how privacy may be at stake. Due to the massive global penetration of services on the Internet and the growing amount of data within it, this in turn requires techniques for abstracting and filtering data without losing information needed for the run-time analysis. In addition, the rapid evolvement of enabling technologies in the cyber space, such as universal connectivity, ubiquity, Cloud services and IoT, introduces a larger attack surface, as well as numerous new threats, attacks and vulnerabilities that know no borders. A challenge is the highly distributed nature of information systems where privacy breaches may be due to the combined information gathered from completely different sources.

2.2. In the Cloud

Cloud providers need to commit to being responsible stewards of other people's information. This implies demonstrating both willingness and capacity for such stewardship. An accountable cloud provider must define what it does, monitor how it acts, remedy any discrepancies between the definition of what should occur and what is actually occurring, and explain as well as justify any action. To achieve this, further research is required on how to enable accountability in the cloud provider chain.

Hosting secured and trusted services or applications in clouds poses new challenges both for the service owners and the cloud operators. Cloud infrastructures introduce new characteristics like multi-tenancy, elasticity, automatic scalability or native distributed

architectures; it departs from monolithic architectures with single points of control where security enforcement and privacy policies can be located. Traditional perimeter security and access control to resources don't cope well with the new level of dynamicity and complexity achieved. Like other processes such as deployment (in a DevOps way), security must now rely on advanced automation and "programmability" to be natively included in all processes from design to operations. Preserving privacy in public clouds or multi-cloud environments is also a new challenge as risks and responsibilities may be spread among multiple actors.

- A specific area for research is to evaluate how cloud providers currently make transparency-related information available to cloud customers, and to assess what are the effects of having transparent services, in terms of costs and benefits to cloud customers and providers.
- Research is also necessary to evaluate how existing or future privacy enhancing technologies (PETs) can be extended for applications in cloud environments or which new requirements for PETs arise from such environments.
- Another challenge is automated security and privacy enforcement and traceability in federated clouds. Distributed applications relying on the integration of multiple services pose new constraints on the interfaces and data exchanged; in addition to the functional description of the services, non-functional characteristics must be formalized, integrated in a consistent framework to match the business requirements and be implemented, tested and maintained in a robust and error-free way, relying on monitoring and automation.
- Design patterns for cloud services and for cloud infrastructures need to explicitly address the security requirements, both at the resources' level and at the composition level, in a consistent constructive framework able to provide the trustworthiness and traceability of these requirements, in the design and build phases, but also when the systems are up and running. Thus, security-by-design in cloud services and infrastructure is a prerequisite.
- Cryptographic mechanisms for data stored and computations performed in clouds or on other service platforms is a research challenge. Cryptography technologies can give the service owners the ultimate control over the privacy and access of their assets. In cloud and distributed environments, advanced mechanisms like blind storage, homomorphic cryptography, advanced credentials delegation and data sharing architectures need to be developed and standardised.
- Known technologies for authentication and access control (e.g., PKI) need to be natively included in all architectures and interfaces supported by cloud systems and maybe adapted to meet the new challenges (scale, number of actors, intensive use for all transactions, ...), so that generalized and standardized identification, authentication and access control is in place.
- The supervision of large services based systems requires new level of cooperation's between multiple actors: network or infrastructure providers, third parties service providers, IT and business managers, users, authorities, etc. Monitoring this kind of system and reacting in real-time to new threats (DoS, intrusions, fraud, identity theft, etc…) must be based on advanced technical components but also in standardized frameworks to ease cooperation and crisis management.

2.3. While Analysing Data

Big Data analytics is often dealing with data coming from different sources and are of different origin. On the one hand this allows more detailed analysis with less dependency on a single data source, but on the other hand it boosts the demand to develop techniques to protect privacy. Traditional security and privacy mechanisms suited for small-scale, static and

localized data are not applicable anymore. Consequently techniques have been developed to prevent leakage of data and to measure the risk of disclosure to scale. Specific research effort has been conducted on privacy-preserving record linkage (PPRL) in the last few years, however there are still challenges privacy poses to data integration in the era of Big Data and analytics.

The following research topics play a vital role in the acquisition, processing, and analysis of Big Data in multiple domains such as healthcare, internet of things, communication, transportation and logistics.

- Privacy related to data processing in clouds, especially if portions of personal data are distributed over several information processing systems and if an unwanted recombination of these data portions would expose a complete set of personal identifiable information is still a research area. Prevention of re-identification of personal information and/or data is still a topic to be evaluated.
- Statistical privacy would be another research area which focuses on restrictions how to process stored data while eliminating all personally identifiable aspects from this data. Its goal is to find algorithms that process data according to these restrictions.
- Data bases such as NoSQL put their priority on performance, thus new security and privacy mechanisms have to keep the data layer thin in order not to impair performance too much. Approaches include offloading the security and privacy functionality onto frameworks, e.g. Hadoop. Here additional research is required to implement such measures directly to the core of the system.
- The risk of re-identification of persons rises with the number of information items (“attributes”), regardless whether or not these items are already personally identifiable information (PII).
- Analytics is applied for a specific purpose or for aiming to get more valuable insights or information needed for specific services. As a result of an analytical process, not all information items are needed for a specific purpose or service, and can be deleted and excluded from further data provisioning processes. In this way, for each information element kept for processing, the purpose or service supported by this element can be given. This requires that analytical results also control the data provisioning process for consequent minimization of collected and stored data (Purpose-driven data minimization).
- In many cases, the collector of the data (data controller) requires support from third parties (data processor) to perform the analytics on this data, also named as Third-Party-Data-Analytics. On top of that, use-cases might require identifying natural persons applying the results of these analytics in order to perform targeted actions. Best practice in this setup would be that the data processor cannot correlate the dataset provided by the data controller to natural persons. Therefore, anonymization techniques must be found, which can only be reverted by the data controller and at the same time which do not make the resulting data unusable for performing the analytics.
- The methods of Big Data and privacy preserving both rely on attributes and insights about data, in some cases the requirements are orthogonal, but sometimes contradicting. A categorization and formulation of principles, both per domain and for the integrated domain, would support further discussions.
- Benefit of Big Data analysis is utilization of several information sources. Today IP addresses or user names are connecting identifiers; both of them inherit privacy issues. Research on further insights about linking parameters is required.
- The dimension of time is a little researched property in the area of privacy protection. Next to research questions such as:
    - What time after an action taken place may knowledge about it be considered non-intrusive?
    - What categories of analytic results shall never be considered non-intrusive?

- Which time frames likely allow for benefit out of Big Data analysis without threat to privacy?

These and many other such questions should be further investigated.

- Today manifold technologies and products exist to detect security incidents and violations; little exists in the area of privacy violations. These tools and technologies to detect privacy violation could be located in the internal (self-governance) or external (public operated) domain, including as well the technical and administrational process.

2.4. Awareness in Life Cycle

The Future Internet will provide an environment in which a diverse range of services is offered by multiple suppliers, often in service-chains, and users are likely to unknowingly invoke underlying services in a dynamic and ad hoc manner. Moving from today's static services, we will see service consumers that transparently mix and match service components depending on service availability, quality, price, and security attributes. Thus, there is the need for service platforms that build the foundation of a Future Internet which is responsive to market needs, but at the same time preserves the security of services and privacy of users' data. Today's, approaches to check systems during design-time are inapplicable for dynamically composed services given that

(i) a service behaves as a black-box, thus a composite service's building blocks are not necessarily open to inspection;
(ii) at runtime, service components may be replaced by new ones; and
(iii) the interaction among different services leads to non-deterministic outcomes.

Security-by-design and privacy-by-design are not sufficient anymore given the above mentioned behaviour. New approaches for software development are pointing to security and privacy by runtime. Security and privacy engineering should be an integral part of software engineering, leading to guidelines, examples and demonstration of the integration and implementation of security and privacy techniques into the actual software engineering flow, from requirements engineering, through design and implementation stages to verification (including proper metrics), assessment, auditing, operation and pre-emptive maintenance. Security and privacy cannot be treated as an add-on but have to receive similar attention as the functional requirements to be achieved.

- Security and privacy considerations must be addressed in all software development lifecycle (SDLC) phases. In the design phase a clear understanding of the privacy and security risks that are associated with a particular service must be developed and appropriate security mechanisms and privacy-enhancing technologies (PETs) must be selected. An important aspect of the privacy-by-design approach is the employment of privacy impact assessments in the SDLC to assess the compliance of the software design to the data protection legislation and social norms and obligations of the service and application providers.
- Once the design is implemented, it must be assured that the implementation really guarantees the security and privacy requirements. Since security is known to be non-compositional, this is a non-trivial task. In particular, when using services, the environment in which the service operates may change. Therefore, the software should also be monitored at runtime for any security or privacy breaches. To this end, security and privacy contracts could be defined that can be monitored, analysed and audited at runtime.
- Data should be equipped with security and privacy policies that have to be obeyed even when the data are passed on to a different processor. In presence of such a sticky policy, a privacy-enabled database should not behave like a store-and-retrieve-

only collector of data, but should also have implemented policies and methods for clearly defined automatic deletion of data. Similarly, security-by-design must take into account that persons may enter malicious data into any information processing flow and be prepared to handle this in a secure and predictable way.

- Privacy-enabled applications must be “certified” or “labelled” to existing regulations and legal requirements. It is necessary to develop novel methods and tools that help organisations to determine whether an application fulfils such standards for information security. In the context of privacy it is about to specify and advertise privacy policies of applications and services similar to the security policies of web services.
- Information security and privacy engineering should not only be treated as a technical discipline but in a multi-disciplinary way. We can observe on the one hand, that there are economic incentives. Well-defined, cost-effective and multilateral procedures and methodologies to engineer security and privacy in new software, systems and services are required. On the other hand, to protect citizens against abuse of their data, the software’s protection mechanisms must be of a high usability, and the software must be suitable to raise the user’s awareness of security and privacy threats. These issues should be investigated in collaboration with psychologists. A third discipline that should also be involved in the effort to provide the society with secure and privacy-aware software is sociology, because software has an increasing influence on the daily lives of huge numbers of persons and their inter-personal interactions.
- Trying to amalgamate security initiatives into an efficient well-built application can be an overwhelming task. There are very few security engineering practices developed specifically for agile processes [7] [8]. Most existing security engineering processes and models, such as Microsoft Security Development Lifecycle[1], Cigital Touchpoints, Common Criteria and standards such as the ISO/IEC 27034, are all based on a traditional, prescriptive waterfall approach [9]. Even measurement models such as the Building Security in Maturity Model (BSIMM) seem to fit better with a traditional non-agile development approach. However, further research is required since agile development does not fit the sequential use of activities or the requirements for extensive documentation of these traditional security engineering processes [10].
- The suitability of traditional security engineering processes has rarely been empirically evaluated in industrial agile development settings [11]. Current research focuses on theoretical analysis of whether certain security practices can adhere to agile principles, concluding that some activities are suitable for agile development and some are not. Thus, there is little empirical evidence on how to implement security practices in agile software development. Besides, there is a need to better understand how to apply security practices across different industrial contexts [12].
- Privacy preserving/enhancing technologies

During the last decades researchers have produced a wealth of mechanisms and tools to protect the privacy of individuals, which have been grouped under the privacy preserving/enhancing technologies term [13]. These technologies may be applied on both the client and the server side of an application, and they can even be deployed as standalone solutions that contribute to protect a data-subject's privacy e.g. the Tor anonymity network [14].

The best known set of privacy technologies are those trying to achieve personal data minimization (i.e. preventing disclosure of identity and personal attributes), though many

[1] Microsoft has introduced an SDL variant for agile development (https://www.microsoft.com/en-us/SDL/Discover/sdlagile.aspx), but there are indications that it is still not agile enough.

others have been developed to support users in managing their personal data, and understanding and developing their privacy practices [15] [16] [17].

Different technologies have been developed to minimize the information disclosed while interacting with other users or services. The communication payloads can be encrypted to render them unintelligible to eavesdroppers, and the state of the practice reports technologies range from basic client-server encryption (e.g. by means of TLS or SSH) to more advanced end-to-end encryption (e.g. using PGP) that makes the content unintelligible even to the service provider. Anonymous communications further protects the identities of the source and destination of the communication by applying techniques such as onion routing or mix-networks.

Other technologies address the concerns related to the publication and mining of information already collected so as to preserve the privacy of the data subjects while maintaining its utility. The state of the art reports different obfuscation and cryptographic techniques to protect privacy when publishing and mining these data sets. These techniques modify the original data yet allow analysis to obtain useful insight. The development of searchable encryption schemes even allows storing of data in remote servers, e.g. in cloud providers, and searching over the encrypted data. Further, privacy-preserving computation methods enable carrying out different computations over cyphered data either by a single entity or by several e.g. somewhat or fully homomorphic encryption, or secure multi-party computation.

Finally, empowering users to exercise control about their data is the goal of some other techniques. Some examples are privacy dashboards, privacy icons, or standalone tools supporting their users in learning and applying common settings in social applications, by default.

While the previous paragraphs provided an overview of the state of the art, there are some challenges that need to be tackled before privacy technologies can become useful for the privacy engineer.

- Further research is required to better communicate information and help users understand the privacy consequences of their online actions, for most users this is difficult, and thus new tools to support end-users in taking decisions are required, such as privacy seals, which may help to reduce the imbalance between subjects and data controllers.
- Basic cryptographic primitives for searching and operating on encrypted data need to be further developed, so as to deliver new functionality [18]. Further, most of these technologies require trading off against other concerns such as usability, performance, scalability, availability, and even utility. For example, some obfuscation-based techniques require trading off privacy against utility, to the point of making them nearly useless. These trade-offs should be further investigated and documented, so as to support engineers in their decisions. Some of these technologies are still academic, proof of concept level implementations that should be taken nearer to the market needs.
- Research is needed on objectives, metrics and evaluation criteria to assess the privacy guarantees provided by different techniques, on their own and when composed with others. In this regard, composability of privacy properties is still a research issue and further results need to be provided on whether when composing several components the privacy properties of the individual components are kept in the larger system.

2.5. Data Analytics for Security and Privacy

Data providers and data processors are responsible to safeguard the information they poses. New technologies and practices (e.g. cloud hosting, distributed services) enable outsourcing

much of the data hosting and processing and seem to reduce the control providers have over their own data. In many cases the adoption of cloud and services is complicated by the existence of legacy systems of records, existing applications and, data bases as well as the need to adhere to regulations. To date, work has been done in the area of data security which enables providers to secure not only data at rest but also data in motion (e.g. secure data moving to web application on the cloud) providing limited solutions for specific cases. Additional research should be conducted to widen the applicability of data security solutions and support a wider range of constraints (legacy systems) as well as deploying existing solutions in real world cases.

Additional challenges lie in the area of privacy, and in particular in the technologies related to publication and mining of information. Clearly, releasing (publishing or mining) data can benefit society (e.g. open data). However, methods must be applied that minimize the risks for individuals [19]. While many privacy preserving methods exist they are lacking in two major areas: Often they may not be able to handle big data and/or unstructured data and in many cases they rely on human input to identify what data is private and/or sensitive. To this, one should emphasise on the fact that data analytics should consider on the location of the data streams that integrate into a data analytics process. Clearly, with the current trends of big data these issues must be addressed.

- Data analytics can be used to support the user with decisions regarding which data is sensitive and to what degree it is sensitive. Automated identification and classification of sensitive information would be able to support users when data sets are very large (big data) as well as when the data itself is unstructured. Data Analysis may also be applied on the results of data mining (e.g. statistics) and provide the user with additional guidance (see section 2.3).
- With the help of Big Data analytics software it is sometimes possible to obtain personally identifiable information by combining different data sources which do not provide such information individually. Research should be focused on the feasibility to use analytics software as a kind of “early warning system” to detect whether data about a user which is distributed over several databases could be misused to reconstruct personally identifiable information.
- Improve existing privacy technologies (e.g. anonymization) such that they become useful and the trade off against other concerns such as usability, performance, scalability, availability, and even utility are diminished (see section 2.5).
- Apply analytics to automatically detect sensitive information in unstructured data, e.g. detect sensitive text in emails. Once sensitive text is discovered a mitigation action should be selected (e.g. masking of sensitive text).
- Exploit analytics to detect privacy and data protection related violations, such as the transfer of personal and sensitive data in authorised locations.
- Often, there are constraints imposed on the data. These may be imposed by an existing DB schema or application API. For example, a DB may expect a phone number to be 7-10 digits. Any solution modifying the sensitive information must adhere to these requirements. Analysis can be designed to identify the set of requirement on the Data. These requirements may include: format (e.g. email), valid values as well and referential integrity requirements. Another example is to select the most appropriate masking scheme (e.g. encryption, redaction, tokenization etc.).

3. Application Sector Examples

Often services applied in application sectors are composed from numerous single building blocks which may be distributed not only to different data centres, but also to different locations, companies, and countries which may have different regulations for data security and privacy. Data is exchanged between these locations using public networks, which may lead to potential data leaks. In the following there are some examples for application sectors where NESSI members are active.

Communication

Communication service provider (CSP) possess huge amount of data. Given that CSPs are acting in a strongly regulated environment business intelligence technologies are packaged to meet the specific sector requirements. Use cases are reaching from churn prediction, customer satisfaction, improvement of network quality, increasing sales, and fraud protection up to operational costs. The main challenges from a technical point of view are (near) real time and predictive analytics, data mining, deep learning and text analytics for massive data. Next to other cases analytics of both structured and unstructured data is required to get deeper insights of subscriber behaviour, their preferences and usage patterns. This may lead to real-time contextual offers, thus achieving real subscriber satisfaction.
Privacy and security are essential in this strongly regulated environment. The exchange, distribution and recombination of data should never violate the privacy rules the CSP claims to comply, thus clear and comprehensive information on how personal data is handled is key.

Internet of Things

The Internet of Things (IoT) connects the physical and the virtual world, offering enormous opportunities for business areas such as industrial manufacturing, predictive maintenance, robotics, smart physical environments, personalised services, smart buildings, command and control and many more. It also introduces additional risks, caused by the sensitivity of data being processed and the increased attack surface. While we assume that the capabilities of devices will increase over time, the major challenge lies in the number of connected devices and the complexity of the resulting system structures, which render traditional security mechanisms (for instance, for distributing security credentials) infeasible. We see and address the major research challenges in identity and key management, the support of security protocols for energy constrained devices, the automation of device on-boarding and maintenance, as well as in the provision of end-to-end security solutions. The development of Privacy-enhanced technologies for data sharing will be of crucial importance.

Web Application Security

The significant part of today's applications is consumed by means of web applications, where active content is consumed by a client via a web browser. Web applications are prone to a large number of vulnerabilities, including those that are known for years as well as those introduced by new web technologies. It is notoriously difficult to spot such vulnerabilities, and application developers are overstrained with understanding and identifying them in their code. There has been considerable investment in establishing state of the art technology for code analysis and security testing as part of the development routine. Research is conducted into increased automation of the approaches as well as the combination of complementary techniques (e.g., static and dynamic code analysis) in order to increase the precision of the results and the efficiency of the analysis.

Application Security Monitoring

In order to cope with the complex and distributed nature of targeted attacks like Advanced Persistent Threats (APT), Security information and event management (SIEM) systems need to extend to the application level by understanding the application semantics and including application logs. As a result, an application security monitoring solution has to cope with very

large amounts of data. Enterprise threat detection solutions combine fast event stream processing with the capabilities of in-memory databases to be able to cope with the data generated by complex landscapes in real-time. The analytical capabilities of an in-memory database allow complementing pattern-based identification of known types of attacks with anomaly search for previously unknown threats.

Application security monitoring can benefit from information sharing, both vertically (across layers) and horizontally (across instances). Since both raw data and analysis results are highly sensitive, information sharing requires proper sanitisation of the data while maintaining its utility for analysis. The research challenge is to optimise the trade-off in order to encourage information sharing while simultaneously respect the protection needs of the data providers.

Critical Infrastructure Security

Industrial Control Systems (ICSs) are vital for the well-functioning of production and distribution of electricity, water treatment, rail signal control, hospital systems, and many other critical infrastructures on which the modern societies depend. ICSs are the "brain" of the critical Infrastructures as they monitor and manage these industrial processes. Most ICSs consist of supervisory software installed on a network of servers, which acquire real-time data from remote devices that control local operations. These supervisory SW collects information about the process and environment, processes the data, displays it in aggregated form to the operators in a control centre and sends back automated or operator-driven supervisory commands to the devices.

ICSs have experienced an abrupt increase in cyber threats and attacks in the last decade and have aroused the interest of security experts. The number of detected vulnerabilities in SCADA systems has exploded. Many critical infrastructures require unattended controllers in the field, which are subject to the physical access of possible attackers. Therefore the security testing and the forensic requirements of embedded devices and applications of ICS systems will play a singular role.

Smart Grids / Smart Meters

The Advanced Metering Infrastructure (AMI), and in particular the smart energy meters will be the initial cornerstone of smart grids. Frequent remote metering of energy consumption (electricity, gas, remote heat, etc.) provides a near real-time feedback into the grid’s load controllers and users, which can take responsive actions to balance demand and response. From a privacy and security perspective the smart meter introduces several worries. An important one is related to consumer trust. While traditional meters offer operators and consumers a relatively high assurance about the accuracy of the readings the new SW-based device placed inside homes could, at least in principle, be prone to cyber-attacks. The smart meter introduces remote connectivity which is not under the consumer’s control and both wireless access technologies and power line communication inside the home may not be under the full control of the operators or customers. Since the demand response control requires detailed usage data and this data provide potentially valuable information about a household’s activities, privacy protection is at stake.

Smart Cities / Smart Buildings

The population of the world’s cities is growing over-proportionally. In 2050, the number of people living in cities will be almost as large as the world’s entire population today. Faced with enormous challenges to be environmentally sustainable and to offer better quality of life to their residents, modern cities are seeking for smart, ICT-based solutions to provide the required smart integration of physical infrastructure and information and knowledge in the areas of education, health care, social programmes, transport, local government and urban planning and management of public services, smarter buildings, and environmental aspects, including handling of waste, pollution, etc.

The smart city will largely depend on automatic information collection from the field, but the availability, resilience and accuracy of these sensor networks must be assured. The transition to interconnected, all-IP networks is a main trend in this area, making cyber security and resilience a key concern, as the failure or impaired performance of the system could disrupt or degrade the operation of the city. The collection and use of information will require sharing and integration of sensor data with data from other city resources creating large smart city datasets. The risk that they could be mined to unveil personal information creates a strong demand for ensuring the personal right to privacy. The challenge is to understand how data can be safely collected, stored, processed in a transparent way and used without citizens fearing their privacy being compromised.

4. Conclusions

Europe should strengthen research and innovation activities in the field of security and privacy to meet the challenges imposed by intensified use of cloud and data technologies. New technologies have to be explored while existing one should be brought into use. Especially from a perspective of Software, Services, Cloud and Data security and privacy is of essential relevance and since the business value creation goes up the stack in most sectors, closer to both the usage of services and users them self, these non-functional properties will be a game changer for business opportunities in the cyber space.

This white paper has identified research opportunities in different sectors, covering Software, Services, Cloud and Data of all kinds where security and privacy are key concerns. Awareness and empowerment of users, compliance with regulations, policies and security and privacy at run-time are some of the essential goals.

The evolution of existing and the development of new technologies will open up a range of novel opportunities, as well as future research and innovation challenges. Joint research and innovation efforts are required and Europe should invest in research and innovation programmes if Europe wants to meet and leverage the opportunities of future ICT trends.

Contributors

- José María del Álamo, UPM
- David Bernstein, IBM
- Pascal Bisson, Thales
- Jorge Cuellar, Siemens AG
- Ariel Farkash, IBM
- Jorge López Hernández-Ardieta, Indra
- Maritta Heisel, paluno – University of Duisburg-Essen
- Martin Gilje Jaatun, SINTEF
- Eva Jaho, ATC
- Volkmar Lotz, SAP
- Roberto Martínez, UPM
- Per Håkon Meland, SINTEF
- Andreas Metzger, paluno – University of Duisburg-Essen
- Micha Moffie, IBM
- Bassem Nasser, IT Innovation
- Eldfrid Øvstedal, SINTEF
- Mika Rautila, VTT
- Valére Robin, Orange
- Nikos Sarris, ATC
- Robert Seidl, Nokia (editor)
- Bjørn Skjellaug, SINTEF
- Daniela Soares Cruzes, SINTEF
- Bjørnar Solhaug, SINTEF
- Mike Surridge, IT Innovation
- Vasilis Tountopoulos, ATC

References

[1] European Convention on Human Rights (ECHR), http://www.coe.int/en/web/human-rights-convention/reference-texts

[2] "REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)". *European Commissioner (January 2012).*

[3] Boehm, B. W., Software Engineering Economics, Prentice-Hall, Englewood Cliffs, NJ, 1981", see e.g. http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20100036670.pdf

[4] European Data Protection Supervisor, Opinion 4/2015, Towards a new digital ethics: Data, dignity and technology, 11 September 2015

[5] European Convention on Human Rights, http://www.echr.coe.int/Documents/Convention_ENG.pdf

[6] EU Data Protection Directive, http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012PC0010

[7] M. G. Jaatun: Hunting for Aardvarks: Can Software Security Be Measured? LNCS Volume 7465, 2012, pp 85-92

[8] L. ben Othmane, P. Angin, H. Weffers, B. Bhargava. Extending the Agile Development Approach to Develop Acceptably Secure Software. IEEE Transactions on Dependable and Secure Computing (TDSC), 2014

[9] D. Baca, Developing Secure Software – in Agile Process. Doctoral Dissertation. Blekinge Institute of Technology. 2012.

[10] K. Beznosov and P. Kruchten, 'Towards agile security assurance', Proc. of the 2004 workshop on New security paradigms, 2004, 47–54

[11] T. Ayalew, T. Kidane, B. Carlsson: Identification and Evaluation of Security Activities in Agile Projects. NordSec 2013: 139-153

[12] A. Alnatheer, A. M. Gravell and D. Argles, "Agile Security Issues: A Research Study," ESEM 2010

[13] George Danezis and Seda Gürses. "A critical review of 10 years of privacy technology." Proceedings of Surveillance Cultures: A Global Surveillance Society (2010)

[14] https://www.torproject.org

[15] Seda Gürses. "Can you engineer privacy?" Communications of the ACM 57, no. 8 (2014): 20-23

[16] PRIPARE, D5.1 State of play Report. 2014. Available online at http://prip areproject.eu/wp-content/uploads/2013/11/D5.1.pdf

[17] George Danezis et al. Privacy and Data Protection by Design – from policy to engineering. ENISA, 2014

[18] PRIPARE, D5.3 Recommendations and Research Agenda. 2015

[19] Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques