VALCRI
VISUAL ANALYTICS FOR SENSE-MAKING IN CRIMINAL INTELLIGENCE ANALYSIS

***VALCRI* WHITE PAPER SERIES**
VALCRI-WP-2017-004
1 January 2017

Security and Privacy Technologies in VALCRI

Patrick Aichroth[1] and Sebastian Mann[1], and Rudolf Schreiner[2]

[1]Fraunhofer Institute for Digital Media Technology
Ehrenbergstraße 31
98693 Ilmenau
GERMANY

[2]ObjectSecurity Ltd.
St John's Innovation Centre
Cowley Road
Cambridge CB4 0WS
UNITED KINGDOM

Project Coordinator
Middlesex University London
The Burroughs, Hendon
London NW4 4BT
United Kingdom.

Professor B.L. William Wong
Head, Interaction Design Centre
Faculty of Science and Technology
Email: w.wong@mdx.ac.uk

EUROPEAN COMMISSION

INTENTIONALLY BLANK

ABSTRACT

This paper outlines the goals and major high-level challenges regarding security and privacy within VALCRI, and provides a discussion of the state-of-the-art and VACLRI approaches to overcome respective limitations in the domains of access control, privacy protection, secure logging and model/policy-driven security and privacy.

Keywords

Security, privacy, information system, access control, Model Driven Security, Model Driven Privacy, PET, re-identification analysis, anonamization, pseudonymization, logging, LEA

INTENTIONALLY BLANK

GOALS

In VALCRI, security and privacy play a crucial role. On one hand, the VALCRI system is handling sensitive LEA information which has to be protected from misuse by insiders and outsiders. On the other hand, the system has to be compliant to privacy and data protection regulations.

Security and privacy have significant overlaps, both in terms of concepts and technologies applied. However, they represent different perspectives, and beyond their commonalities, each of them also requires specific methodological, technological and organizational elements.

VALCRI security goals focus on fairly concrete demands related to data integrity and confidentiality, controlling access to structured and unstructured data and avoiding leakage of sensitive data. That includes support for secure collaboration within LEA teams, i.e. to manage access in compliance with the defined rules when users request to share information within their teams, thereby taking the burden of respective manual checking from users. It also includes a protection of the VALCRI system from outside attacks.

Privacy, on the other hand, includes a broad range of context-depending concepts, including e.g., confidentiality, control over data, authenticity, availability, freedom from observation and transparency. The exact interpretation of what privacy means in practice differs among countries, cultures and individuals, at least to some extent. As a consequence, privacy regulations and respective requirements tend to be broader and vaguer than security requirements, and differ across various jurisdictions. Hence, the technological approaches to address them need to be flexible, but they share common themes: The overall goal is to prevent processing of person-related and sensitive personal information where it is not allowed or unnecessary, to support privacy audits, and also to provide means for data removal to avoid that data is stored beyond the legally allowed durations.

MAJOR CHALLENGES

VALCRI is about an evolving, multi-user knowledge system containing potentially very sensitive information. It needs to align the needs for advanced analysis functionalities *and* the equally important need to address security and privacy aspects. If the VALCRI system is unable to meet the legal data protection requirements, it must not be used by police organisations. If the VALCRI system is not able to meet the police organisations' security requirements, they will not use it.

Hence, it is clear that there are difficult security and privacy challenges to be addressed:

(i) Security and privacy approaches should be implemented in the least obtrusive manner possible, which requires e.g. consideration of the context and purpose of analysis tasks, to be able to control access and processing where needed, but to not interfere with legitimate tasks of the analysts

(ii) Security and privacy approaches should be able consider legal differences and individual self-imposed security and privacy constraints per institution

(iii) For certain circumstances, e.g. emergency situations, security and privacy approaches should be able to allow a certain degree of freedom to "override" ex-ante control mechanisms, and apply ex-post control mechanisms instead

(iv) Security and privacy approaches should be implemented in a way that support transparency and auditing from a high-level policy to the actual low-level enforcement, in order to demonstrate that the system is compliant to data protection regulations

(v) Addressing security and privacy requires a variety of technological approaches, which need to be fully integrated in order to exploit their synergies

(vi) The individual security and privacy technologies include many technological challenges, which will be further explained in the following.

To the best of our knowledge, several of the aforementioned problems in the LEA domain have not yet been solved. And while it will not be possible to address all of them within the project, we are confident that VALCRI will provide some security and privacy results which go beyond the state-of-the-art.

In the following, we will outline relevant state-of-the-art approaches for security and privacy, and describe the respective VALCRI approaches.

SECURITY: SOTA AND VALCRI APPROACHES

Data confidentiality and integrity

The main security requirements in the project are **data confidentiality and integrity**. Confidential information must only be disclosed to authorized users, and only authorized users are allow to modify data, in both cases in accordance with the organisation's overall security policy.

This can partially be addressed with standard approaches based on cryptographic techniques, which have to be integrated within the data ingestion work flows as well as structured and unstructured storage. The main enabler for implementing confidentiality and integrity is access control

In contrast to access control in other systems, **access control** requirements in this project are much more challenging: VALCRI is a multi-user system, which processes highly critical information at different levels, from open information to highly sensitive information about covered actions or data obtained from secret informers.

The disclosure of such information can not only threaten LEA work, it can even cost lives.

In order to prevent this, it is essential to have a sound design regarding access control for the integrated software

prototype that the project delivers: In contrast to other system functionalities, such security and privacy features cannot be provided as add-ons later on. But which access control approaches can make sure that regulatory, legal and operational requirements will be considered?

Access control is a well-established topic in IT security, with Mandatory Access Control (MAC) or Role Based Access Control (RBAC) being the dominant models and the state-of-the-art in practical implementations to protect real-world systems. And at first glance, MAC and RBAC seem adequate for intelligence systems like VALCRI: For MAC, information is tagged with partitions and labels, for example "CaseX, SECRET". A user can only read it with a respective SECRET clearance for this partition, and is unable to write the information to a domain that has a lower clearance than SECRET for this partition e.g. to another user not involved in a case. But while all this sounds sufficient at first sight, practical experience has shown that the approach is problematic, because it becomes unmanageable within larger environments. Not even very security sensitive organisation are able to run multi-level systems, which means that their systems are mostly used using a one-clearance approach allowing legitimate users to access (almost) all information. As a result, there are frequent cases of disclosure, which as we have mentioned above, are extremely problematic. In addition, it also limits the exchange of information, which can lead to major operational risks.

Hence, in order to enforce fine-grained access control policies derived from regulatory, legal and operational requirements, it makes sense to use more recent access control models, such as Attribute Based Access Control (ABAC) and Proximity Based Access Control (PBAC). ABAC allows deriving access control decisions not only from labels attached to data, but also from attributes of structured data. For instance, given that the system keeps provenance information which can identify the original source of data (used in the analytical process to establish trust into data), ABAC can also use such information attributes for the access control decision. For instance, if provenance data signals that the information originates from an informer, then access may only be granted to the data owner and specifically authorized users. Similarly, if data is bound to a specific purpose, it can only be accessed if an analyst is working in the context of this purpose, e.g. a specific case. Other access attempts will be denied.

PBAC extends the concept of ABAC, allowing access decisions based on a generalized concept of distance. The concept of distance is not limited to a geometrical or geographical distance, but can also include e.g. hierarchies or social distance. PBAC allows policies like “access all records in an area of 5km around a place” (geographical distance), “the supervisor is allowed to see all records of the analysts she is responsible for” (hierarchical distance), or “allow access to all direct contacts of a suspect” (social distance). Moreover, it is of course possible to use different attributes, and use ABAC and PBAC terms in one access control rule, so e.g. the terms “no data from secret informers” and “only in a specific area” could be combined.

In combination, ABAC and PBAC are capable to define and enforce fine-grained access control polices based driven by legal, regulatory and operational requirements. However, these powerful models are very difficult to apply in practice due to their complexity: In many cases, security policies are expressed as attributes which are not directly accessible from underlying security mechanisms or the functional system. Instead, they have to be transformed from other attributes. For example, it may not be possible to directly query whether a user is from a Schengen state, but instead, the user identity is mapped to an organisation, the organisation is mapped to the respective country, and then it is possible to check whether this county is member of the Schengen Agreement. Similarly, a geographical position to be used for an access control decision could be obtained in different ways: The user (or a user's device) provides the position, a third party system like a cellular communications network provides the position or the position is somehow stored in the system, e.g. as part of a job responsibility. Different sources of attributes enjoy different level of trust in the content of the attribute, and analysing the different attribute chains and levels of trusts is almost impossible for humans.

Model-driven security and OpenPMF

In order to overcome the aforementioned issue of complexity, the innovative VALCRI approach is based on an implementation of **Model Driven Security (MDS)** concepts for ABAC and PBAC. MDS and its implementation, ObjectSecurity's **OpenPMF** Policy Management framework, were originally developed to define and enforce correct security policies for complex distributed systems. In VALCRI, MDS concepts and OpenPMF are extended to handle complex attributes and rules based on them:[1] MDS/OpenPMF with support for ABAC and PBAC allow to express and enforce fine-grained, high- level security policies driven by legal, regulatory and operational requirements, to generate human readable representations of the policies for human audit, to test the policies, and to enforce them in a complex distributed system.This is how complexity is being dealt with for access control, but this also provides a key element to other aspects, as we will see in the following.

There is a highly challenging **gap between high-level security and privacy requirements and low-level enforcement**:

As experience has shown, humans are not able to translate high level security policies based on operational, regulatory or legal terms to low level security enforcement rules and configurations in complex systems. This is caused by the high complexity of today's systems. Humans do not

[1] U. Lang and R. Schreiner. Proximity-Based Access Control (PBAC) using Model-Driven Security, ISSE 2015

understand all interactions between the system's components and the required protection mechanisms in sufficient details and accuracy. They are not able to write long access control rule files in languages like XACML. In addition, the produced configuration is not verifiable, there is little assurance that the security mechanism really implements the high level policy.

In order to address that, VALCRI implementsn iterative approach where both high-level requirements and low-level technology components are formally captured and mapped to each other using a combination of top-down and bottom-up approaches: A Domain Specific Language is first used to precisely capture high-level policies, and model transformations translate these policies into the matching technical enforcement rules and configurations. Afterwards, low-level technical components are matched to the proposed specifications, to derive which of the requirements are actually technically implementable and addressing which high-level goals. The desired result of this iterative process is to get to a technical implementation methodology that implements all the key requirements. In addition, a human readable documentation of both high level security policies and low level implementation rules are also generated.

While this approach has been intended for access control (which is the key technology required to implement many security and privacy requirements), it can at least to some extent also be applied to e.g. Privacy-Enhancing Technologies (PET). By doing do, the MDS/MDP-based VALCRI approach is effectively realizing a new, integrated policy-driven approach to several security and privacy technologies. As a result, it is possible to address several of the key challenges mentioned at the beginning of the document: (a) legal differences and individual self-imposed security and privacy constraints per institution can be considered by defining different policies, (b) auditing is supported, and transparency regarding policy definition promoted and (c) synergies among very different technologies can be exploited.

Secure logging

Finally, a last key ingredient to the mix of security and privacy technologies is **secure logging** in order to support ex-post auditing and examination of whether actions were appropriate and effective. In intelligence, it is crucial to allow a certain degree of freedom to "override" ex-ante control mechanisms, and apply ex-post control mechanisms instead: In many circumstances, it is difficult to determine and enforce beforehand what is necessary and proportional, this especially goes for (but is not limited to) privacy constraints. In such cases, logging provides a means to "trade" less ex-ante control for more ex-post control, allowing a detailed examination of actions and decisions afterwards.

Secure logging requires that only authorized personnel / auditors should be able to access logs. Moreover, logging should be kept completely separated from the operational system, allowing read access only to avoid manipulation. Here, there is an obvious challenge: Who guards the guards? A system administrator may not only able to misuse information, but is also able to cover traces by deleting the related log files. In order to address this challenge, we use a mainstream log system for the collection, storage and analysis of the log data, but we add an additional element: **High Assurance Logging and Auditing (HALA)**: By using an additional hardware device that implements a "Vault", a separated high assurance domain that cannot be accessed by the system administrator. This device ensures that even the system administrator is not able to delete files or log files without leaving traces. Appropriate authorities, e.g. audit officers, can then always check whether log files were deleted or modified.

PRIVACY: SOTA AND VALCRI APPROACHES

In the following, we will outline relevant state-of-the-art approaches for privacy protection, and describe the respective VALCRI approaches.

Face encryption for videos

There are different approaches for blurring faces in videos to preserve the privacy of the people caught on camera. For a certain time this was the preferred method to obfuscate faces in images and videos, but there are some new algorithms to perform face recognition even in blurred images/videos[2].

In order to address this development, it is planned to also integrate **face encryption for videos**: We have implemented components for video face detection and respective h.264 video encryption (which is a significant challenge by itself, especially due to the need to implement in-place encryption of video regions within the encoded video stream). Face encryption in VALCRI is done by applying face detection (developed by Fraunhofer IIS) to detect and track face regions. Based on this information, the detected regions are encrypted using a standard AES256 algorithm, applying a modified h.264 encoding process that allows selective encryption. Each face is encrypted using an individual key (stored together with its position in an encrypted XML-file), which allows users to selectively decrypt specific faces for a specific time period, while leaving other faces encrypted.

In order to integrate face encryption with the MDS/MDP-based approach, it is now necessary to adapt OpenPMF to the spatio-temporal video fragment support, and implement respective low-level enforcement.

PET

Regarding **Privacy-Enhancing Technologies (PET)**, VALCRI however includes more than face encryption. One

[2] McPherson, R., Shokri, R., Shmatikov, V.; "Defeating Image Obfuscation with Deep Learning", see https://arxiv.org/pdf/1609.00408v2.pdf

goal of PET in VALCRI is in providing components for data anonymization and pseudonymization. As for access control, the goal is to apply such technologies in the least obtrusive as possible, preventing unlawful processing of person-related and sensitive information, and to use PET to generalize data where it could otherwise not be used, or would violate privacy unnecessarily.

This is achieved by implementing an **anonymization / pseudonymization toolset** that includes state-of-the-art algorithms e.g. for randomization and generalization[3]: *Randomization* means adding a random element to the data. This could be done using permutation (shuffling data elements), adding some noise to the data, or by blocking sensitive data. *Generalization* means to reduce the codomain of the data elements, e.g. by generalization of values (e.g. $10 \leq \text{age} \leq 19$, instead of age = 18), aggregation of values (using larger areas like counties, instead of specific cities) or methods like k-anonymity and alike.

Instead of applying such tools in a static way, however, we ensure their effectiveness by dynamically parametrizing them using a 2nd toolset for **re-identification analysis**: A common problem for anonymization and pseudonymization is that data combinations often allow identification of persons based on data that, by itself, would not allow identification of persons. Even worse, it is difficult to even find tools that measure such re-identification risks. In order to address that, VALCRI includes development of re-identification analysis components: By using statistical analysis, we do not only measure the risk of person identification, which allows LEA to e.g. process data before it is exported. We also use the results to parametrize the aforementioned anonymization / pseudonymization tools, making them more effective and suitable for data import and export, avoiding a lot of manual work, and introducing an additional tool to ensure legal compliance for data access based on the respective policy.

Combining security and privacy technologies

The VALCRI approach regarding security and privacy is different in that both the security and the privacy toolsets are integrated into the MDS/MDP-based approach outlined above, allowing for a combination of access control and anonymization & pseudonymization.

This combination results in several "access types" beyond a simple "full access" or "no access" which can be used, for instance:

- *partial access to unstructured data*: e.g., videos are accessible but faces (or certain faces) are only visible to authorized users
- *partial access to structured data*: e.g., data is available in principle, but certain attributes (e.g. ethnicity, political views, sexual orientation), or certain data entries are only available upon authorization
- *access to modified data*: e.g. for trend analysis, anonymized data is used, which allows collection and usage of data that would otherwise be impossible to use
- *relevance flagging*: e.g. for very sensitive information, the system signals to the user that there is a statistical anomaly, without revealing the actual data, which requires authorization
- *existence flagging*: e.g. for informer information, the system signals only the existence of data, hinting to a respective authorization process

It is important to note that all the example cases represent *possible* options – whether or not they will be used will be depend on the policy and hence the respective laws and LEA rules. However, they significantly extent the possibilities of how to deal with certain types of data that currently cannot be dealt with properly.

SDB and UDB

The MDS/MDP-based approach requires adapting security and privacy technologies to the all relevant system domains in order to enforce the policy. First and foremost, it requires an implementation of OpenPMF-based access control and PET for the relevant databases. In VALCRI, these are the structured databse (**SDB**, where Fuseki/Jena is used) and unstructured database (**UDB**, where ObjectStore[4] is used) implementations. SDB and UDB require completely different approaches as to how security and privacy low-level enforcement are implemented.

COLLABORATION WITH PARTNERS

All activities related to the development of technologies to address Security, Ethical, Privacy and Legal (SEPL) aspects in VALCRI require a strong collaboration with other partners, and this collaboration, at least so far, has worked very well. In includes:

(a) **interaction with legal partners** (ULD, KUL and MU), in order to determine which legal constraints need to be modelled and enforced, and to define example policies for testing and development purposes.

(b) **interaction with data modelling partners** (LIU) to ensure that all information / attributes necessary to enforce security and privacy are included in the model. This includes e.g. user role, data provenance, type, purpose, etc. but also contextual information, e.g. the notion of urgency. By integrating all related concepts, the model provides the basis to ensure interoperability

[3] EC - OPINION 05/2014 ON ANONYMISATION TECHNIQUES, see: http://www.cnpd.public.lu/de/publications/groupe-art29/wp216_en.pdf

[4] ObjectStore is a Fraunhofer IDMT component that provides access control, ID and other functionalities on top of MongoDB

between the aforementioned security and privacy technologies, but more importantly, it also ensures interoperability with all other system domains

(c) **interaction with key UI/UX and DEA partners** (MU, UKON) to agree on protocols on how security / privacy components interact with UI/UX, and with data analysis and extraction components, especially considering access control and the possibility of data transformations.

CONCLUSION

Security and privacy represent key elements of VALCRI, and the approaches taken in the project go beyond the state-of-the-art especially in the domains of access control, secure logging, face encryption, re-identification analysis, and adaptive anonymization / pseudonymization. Most importantly, and in contrast to many other projects, VALCRI applies a policy-driven approach (MDS/MDP) to combine all of the aforementioned technologies, which allows us to fully exploit their distinct strengths, while keeping the system flexible enough to address very different requirements.

The research leading to the results reported here has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) through Project VALCRI, European Commission Grant Agreement Number FP7-IP-608142, awarded to Middlesex University and partners.

	VALCRI Partners	Country
1	Middlesex University London Professor B.L. William Wong, Project Coordinator Professor Ifan Shepherd, Deputy Project Coordinator	United Kingdom
2	Space Applications Services NV Mr Rani Pinchuck	Belgium
3	Universitat Konstanz Professor Daniel Keim	Germany
4	Linkopings Universitet Professor Henrik Eriksson	Sweden
5	City University of London Professor Jason Dykes	United Kingdom
6	Katholieke Universiteit Leuven Professor Frank Verbruggen	Belgium
7	A E Solutions (BI) Limited Dr Rick Adderley	United Kingdom
8	Technische Universitaet Graz Professor Dietrich Albert	Austria
9	Fraunhofer-Gesellschaft Zur Foerderung Der Angewandten Forschung E.V. Mr. Patrick Aichroft	Germany
10	Technische Universitaet Wien Assoc. Prof. Margit Pohl	Austria
11	ObjectSecurity Ltd Mr Rudolf Schriener	United Kingdom
12	Unabhaengiges Landeszentrum fuer Datenschutz Dr Marit Hansen	Germany
13	i-Intelligence Mr Chris Pallaris	Switzerland
14	Exipple Studio SL Mr German Leon	Spain
15	Lokale Politie Antwerpen	Belgium
16	Belgian Federal Police	Belgium
17	West Midlands Police	United Kingdom