FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE

ASSIGNMENT OF BACHELOR'S THESIS

Title:	Analysis of data security from online data-taking to publication in ATLAS at CERN
Student:	Jakub Weisl
Supervisor:	doc. Dr. André Sopczak
Study Programme:	Informatics
Study Branch:	Computer Security and Information technology
Department:	Department of Computer Systems
Validity:	Until the end of summer semester 2021/22

Instructions

The ATLAS experiment at CERN collects data of annual size 30 PetaBytes. This data is created online and direct access is very restricted to a few members of the ATLAS collaboration. Online data access takes place several times per hour and access is currently granted by the control room leader individually. The data is transferred to a CERN storage system and then uploaded to wlcg.web.cern.ch. It is made accessible to participating institutions via virtual organizations to avoid unauthorized access by non-ATLAS members. The rights to access the data remains in the control of the ATLAS collaboration for data analysis and publications.
Tasks:
1) Review the chain of processes regarding data security.
2) Analyse the security of the primary online data access and propose an automation.
3) Investigate possible shortcomings in the data protection on the user side regarding data access and use.
4) Suggest data access improvements for users regarding the control of exploiting the data.

References

Will be provided by the supervisor.

prof. Ing. Pavel Tvrdík, CSc.
Head of Department

doc. RNDr. Ing. Marcel Jiřina, Ph.D.
Dean

Prague January 4, 2021

FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE

Bachelor's thesis

Analysis of data security from online data-taking to publication in ATLAS at CERN

Jakub Weisl

Department of Computer Systems
Supervisor: doc. Dr. André Sopczak

February 14, 2021

Acknowledgements

I would like to thank to all who helped me with this thesis. Namely to my supervisor doc. Dr. André Sopczak, who guided me through this whole process, to Ing. Josef Kokeš, who did not hesitate to help me with theoretical materials and was able to point into right direction and last to Jakub Suchý, for consulting through the practical part.

Declaration

I hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis.

I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No. 121/2000 Coll., the Copyright Act, as amended. In accordance with Article 46 (6) of the Act, I hereby grant a nonexclusive authorization (license) to utilize this thesis, including any and all computer programs incorporated therein or attached thereto and all corresponding documentation (hereinafter collectively referred to as the "Work"), to any and all persons that wish to utilize the Work. Such persons are entitled to use the Work in any way (including for-profit purposes) that does not detract from its value. This authorization is not limited in terms of time, location and quantity. However, all persons that makes use of the above license shall be obliged to grant a license at least in the same scope as defined above with respect to each and every work that is created (wholly or in part) based on the Work, by modifying the Work, by combining the Work with another work, by including the Work in a collection of works or by adapting the Work (including translation), and at the same time make available the source code of such work at least in a way and scope that are comparable to the way and scope in which the source code of the Work is made available.

In In Prague on February 14, 2021 . . . . . .

Czech Technical University in Prague
Faculty of Information Technology

Citation of this thesis

Weisl, Jakub. *Analysis of data security from online data-taking to publication in ATLAS at CERN*. Bachelor's thesis. Czech Technical University in Prague, Faculty of Information Technology, 2021.

Abstrakt

Tato práce se soustředí na dvě témata, první vzdálený přístup k systémům sběru dat z experimentu ATLAS a druhý je ochrana a správa naměřených a nasimulovaných dat v rámci projektu ATLAS. V práci je nejdříve prozkoumán a zhodnocen dnešní stav věcí a na základě nálezů jsou navržena tato řešení. V oblasti vzdáleného přístupu, zamezit přímému přístupu na sytému sběru dat z internetu a přistupovat do sítě experimentu ATLAS pomocí VPN. Změny ve správě uživatelských účtů a přístupových práv na jednotlivé systémy. Posledním návrhem ke vzdálenému přístupu je změna autentizačního mechanismu a použití multifaktorové autentizace. Pro správu dat a jejich ochranu představuje největší riziko možnost stáhnout si data na lokální stanici. Tato možnost musí být však zachována, proto práce navrhuje vývoj vlastního digital rights management řešení, neboť v dnešní době není k dispozici žádné, pro tento formát dat, již hotové řešení. Navrhované řešení se soustředí na znemožnění použití dat bez souhlasu jejich vlastníka (ATLAS collaboration). Toho je docíleno pomocí zašifrování dat, kde dešifrovací klíč je držen ATLASem a je propůjčen na vyžádání. Jako největší problém se v obou případech jeví použití osobních stanic uživatelů, neboť všechna bezpečnostní opatření musejí být dělána na straně ATLASu.

Klíčová slova CERN, ATLAS, informační bezpečnost, přístup, ověření totožnosti, nakládání s daty

Abstract

This thesis focuses on two main objectives. The first is remote access to online data taking systems at ATLAS experiment in CERN. The thesis, after assessment of current situation and recognising areas of possible improvements, suggests the improvements to the ATLAS environment in form of disabling direct access to online data taking systems from the internet and using VPN to access ATLAS network, suggesting different account management and access policies and proposing different authentication scheme where multi factor authentication is used. The second one is data rights management and data control of ATLAS measured and calculated data. System of downloading data to local stations makes really hard to enforce any restrictions. However, it is necessary to keep this option, and so development of data rights management solution is suggested, because no commercial or open source alternatives are available for this format of the data. This data rights management solutions encrypts data which leave ATLAS servers and so request for decryption key must be made towards ATLAS when those data are accessed. Assessment of current situation in both main objectives reveals that the greatest problem is in use of uncontrolled personal stations and so all security precautions must be taken on side of ATLAS.

Keywords CERN, ATLAS, information security, access, authentication, data control

Contents

List of Figures

Introduction

The Large Hadron Collider (LHC) in Switzerland is the most significant project in the field of particle physics in the last decades. It brings together physicist, engineers, IT specialist and many more experts from all over the world to collaborate. This collaboration brought to this world many great inventions of today's world, most significant is probably World Wide Web.

The LHC was constructed and is maintained by the European Laboratory for Particle Physics (CERN). CERN hosts four major experiments ATLAS, ALICE, CMS and LHCb underground on the LHC ring. Each experiment has a different focus. The corresponding collaborations are formed as group of institutes participating on those experiments. CERN is also an participating institute. The experiments vary in the number of collaborating institutions and number of participating physicists. The focus of this thesis will be on the largest experiment the ATLAS.

The ATLAS experiment is operated by the ATLAS Collaboration which brings together approximately 3000 publishing scientist including 1200 doctoral students and many technical staff from currently 181 institutions across the globe.

The ATLAS experiment produces about 3200 terabytes of raw data every data-taking year. These data are the most valuable property of the experiment and therefore it has to be well protected and managed. One of the important task is keeping track of all the copies of the collaboration data. In case an institution or an individual loses the right to access the data, ATLAS must be able to restrict the access to those data across all the copies, even the local ones. Primary data storage is the Worldwide LHC Computing Grid (WLCG), which incorporates computing and data centers from many institutions all around the world. The WLCG is organised in the Tier system where CERN is Tier-0 then there are 13 Tier-1 data centers which each carries identical copy of the data. There are approximately 160 Tier-2 centers which each carries part of the data for local analysis. This solution achieves the desired computing power and the storage, and it requires sophisticated data management.

The ATLAS experiment has many components which are serviced by various teams, and in the whole collaboration there are many people with various tasks. These people, components and different tasks require complex account management. The complexity is even greater when basic information security paradigms are enforced such as least privilege paradigm in form of privilege access management (PAM).

The various security rules and policies effect every member of the collaboration with access to the ATLAS computing system. Therefore, every implemented security policy must take into consideration not just security aspects and sensitivity of certain systems or data but also user comfort because when set up policies starts to be too uncomfortable for users then users try to figure out a way how to bypass. This usually leads to somehow compromising the security. As example we can look at password policies when they are too harsh then users start to write passwords down in better case into their cell phones in the worst case on paper which they stick to the keyboard.

The ATLAS experiment is composed of many computing systems such as the magnet system which bands the path of the elementary particles after the initial proton-proton collision in the LHC, the calorimeters, Muon detectors and other sub-detectors. The ATLAS computing farm takes care of initial data processing and the ATLAS trigger system handles the data selection from all sub-detectors. With rate of 40 millions collisions per second there is a large amount of the raw the data processed and stored. Those data are valuable property of the ATLAS collaboration and must be properly protected.

This raises two main topics addressed in this thesis:

- At first the thesis focuses on managing the data mainly in particular how to protect the data against illegal distribution and publication by authorised users and revoking access to the data for users who have lost the access rights, even to the offline copies which they could have made. This task should be achieved by using data rights management (DRM) techniques and encryption of the data.

- The second topic concentrates on the remote access to the Trigger system during the data-taking process. In particular, on the issue of automated secure remote access to the Trigger system during online data-taking which would mitigate the possibility of human error while granting this access manually.

In the first chapter of this thesis, Theory of information security and guidelines are given. This chapter also gives short overview of the theory about authentication, least privilege paradigm and access management in order to introduce the problematic of securely accessing computing system. Then DRM is described in relation to data access.

The chapter 3 focuses at first on the Worldwide LHC Computing Grid mainly on its architecture, functioning policies and business processes. The

WLCG is a key component because it stores and process all the data taken by the ATLAS experiment. The section ATLAS experiment addresses the whole process of data taking including the trigger system. It is in a close relation to the WLCG because it selects raw data for further processing and storing.

In the chapter Analysis of current situation in information security, the current situation from the information security point of view is assessed and the potential weaknesses are emphasized with main focus on remote access to online data taking systems during data taking and data rights management on ATLAS Computing Grid.

Chapter Suggested solution based on current situation provides suggestions for the design of automated secure remote access to the online data taking systems and suggestions for improving the data rights management of ATLAS data for greater control. The conclusion is given in chapter Conclusion.

Chapter 1

Goals

The main goal of this thesis is to analyze and make suggestions for improving the data security in the ATLAS experiment. This means it impacts all users who access the data, or remotely access the ATLAS systems.

This thesis will suggest possible improvements regarding data management and data access of authorised users. This access will also include full control over all copies of the data.

Another aspect concerns suggestions to improvements for automated access to the Online data taking systems during online data taking for authorised users.

The mentioned suggestions will be tailored for the specific use and architecture of the ATLAS infrastructure and business processes. They will be detailed enough that the ATLAS management could consider their implementation base of this thesis.

Chapter 2

Theory of information security and guidelines

At first term user must be specified. This thesis uses term user as a system user as defined in RFC 4949 "a system entity that consumes a product or service provided by the system or that accesses and employs system resources to produce a product or service of the system".[1]

For purpose of this chapter meaning of word impossible is little broaden. In addition to its original meaning it means action is not possible in acceptable amount of time with today knowledge and technology. This meaning is applied primarily when talked about cryptography and hash functions.

Main goal of cryptography is to restrict access to an information only to subjects for whom the information is intended. This is done by using encryption algorithm which masks and diffuses an information across a whole message. Such algorithm requires a key or keys for its functioning.[2] There are two types of algorithms symmetric-key algorithm which uses one key for encryption and the same or easily derivable key for decryption process and asymmetric-key algorithm which uses pair of keys one for encrypting and one for decryption.

For the encryption algorithms to behave as expected and provide desired classification of an information it is necessary that initial keys must be sufficiently long and randomly generated.

2.1 Symmetric cryptography

Symmetric cryptography uses much less computing power than asymmetric cryptography. This is caused by the mechanism of encryption when individual operations done with plain text are very simple and very fast usually one operation is one or few processor instructions such as several rounds of switching bytes in message. Algorithms for asymmetric cryptography involve in general

more complex steps such us multiplication operations etc.

Use of symmetric cryptography poses one great obstacle which must be overcome every time it is used. This obstacle lies in distribution of a decryption key which is necessary for successful decryption of a message. However fast decryption and decryption process makes them desirable when large amount of data must be handled or data must be encrypted and decrypted with the smallest delay possible. This problem is usually solved by combining symmetric-key and asymmetric-key algorithms together when key for symmetric cryptography is transferred using asymmetric cryptography or by using algorithm for creation of common encryption key such as Diffie-Hellman key exchange.[3]

2.2 Asymmetric cryptography

Asymmetric-key cryptography uses a pair of keys instead of one key. From this pair one key is called public and is given to other users to encrypt data intended for the owner private key which is used for decryption of those data. Those keys are nearly impossible to be deduced one from another. This relies on mathematical theories and problems such as integer factorization problem which used in RSA cipher. The problem is caused that no algorithm for factoring large numbers in known and so it is not solvable in acceptable time.

This thesis will use primarily asymmetric cryptography so its possible uses and principals are explained more detailed below. As was mentioned before asymmetric cryptography uses a pair of keys. One is public key which is publicly known and second one is private and it is necessary to stay secret. To demonstrate how asymmetric cryptography works an example is described below.

To safely communicate user A must encrypt data using public key of user B. To retrieve data user B must decipher the data which was encrypted by user A, by using his private key.[4]

2.2.1 Digital signature

Purpose of digital signature is to prove that an information comes from particular source and that an incoming information have not been altered through the way. This goal is achieved by using asymmetric cryptography in opposite way then in mentioned previous example along with hash functions.

Hash function is unidirectional function that means that it is impossible to to deduce input parameters from a result even though an algorithm is known. A result of hash function is called hash and it is always the same length. Hash function acts as random oracle this means that hash function returns pseudo-random result form defined domain for every unique input however when the same input is repeated than the result must be always the same. Pseudo-random results implies that even little change in input parameters

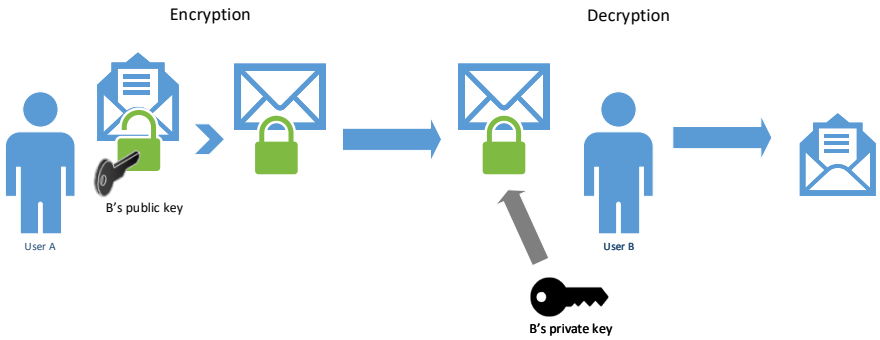

Figure 2.1: Asymmetric cryptography

should lead to completely different result. Some examples of hash functions used today are SHA-256 and SHA-512 where number of at the end hints the length of a resulting hash in bits.

To digitally sing some message an owner of a message first creates hash of the message together with timestamp. This hash is then encrypted using owner's private key (this key should be known only to the owner). This encrypted hash is appended to the message. This message afterwards send to chosen receiver. The receiver of this message deciphers the hash using the owner's public key (receiver already has it, it was appended to the message or is publicly accessible in form of certificate). This part proves an authorship of the message. When public key is appended to the message or gained from public site it should have form of certificate (2.2.2) to ensure that an imposter did not use his pair of public and private key. Receiver then makes hash from the message and compares the two hashes if they match message was not altered during the way.[4]

2.2.2 Digital certificate

Digital certificate is in general a public key and other information which depend on format of a certificate. The most important information which are

always included are data of the owner, a certificate identifier, an expiration date, public key of certification authority (CA) which signed the certificate, and a digital signature of the certificate.

A CA is trusted entity entitled to validate information on certificates and approving them for use by digitally signing a certificate. This raises some security issues when a private key of the CA is exposed because all certificates signed by this authority suddenly becomes untrustworthy. The CA must also administer certificates which have signed, that means it must mark invalid and untrustworthy certificates. That is done by regularly publishing a certificate revocation lists (CRL) which is signed by the CA to ensure its authenticity. This action as well as accepting certificate for signature can be delegated by CA to different systems. Those systems are called registration authority and CRL issuer and they must be trusted by the CA.

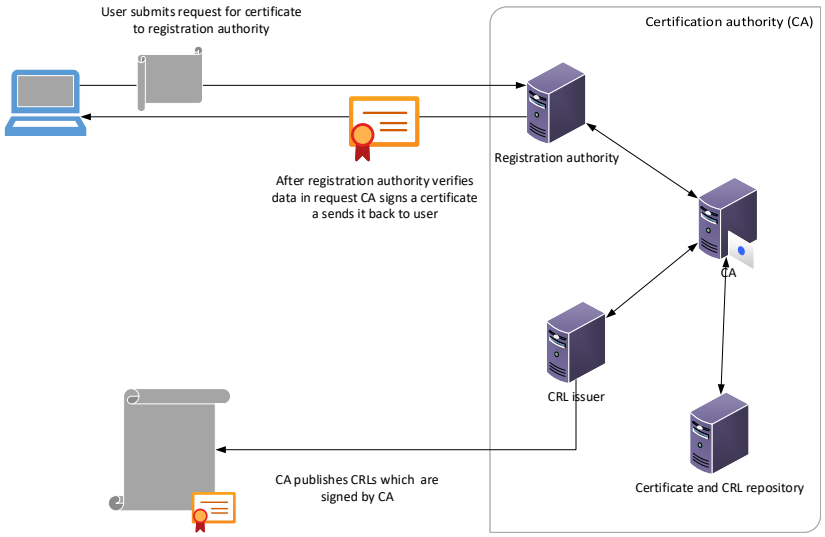

Figure 2.2: Certification authority and process of issuing new certificate

In order to spread the load of requests and make validating data of applicants easier CA can create subordinate CA by signing their certificates and declaring that this CA is trusted by the delegating CA and so it should be trusted by others as well.This trust delegation of CA is called chain of trust.

If any of the subordinate CA or root CA in the chain is compromised all subordinate CAs and certificates issued by the compromised CA or subordinate CAs are compromised as well. This means, that in order for certificate

to be valid, all CA in its chain of trust must be trustworthy. This problem can be avoided with cross signing of a certificate. This is when one certificate is signed by more then one CA however this brings new problems with revocation of such certificates.[5, 4]

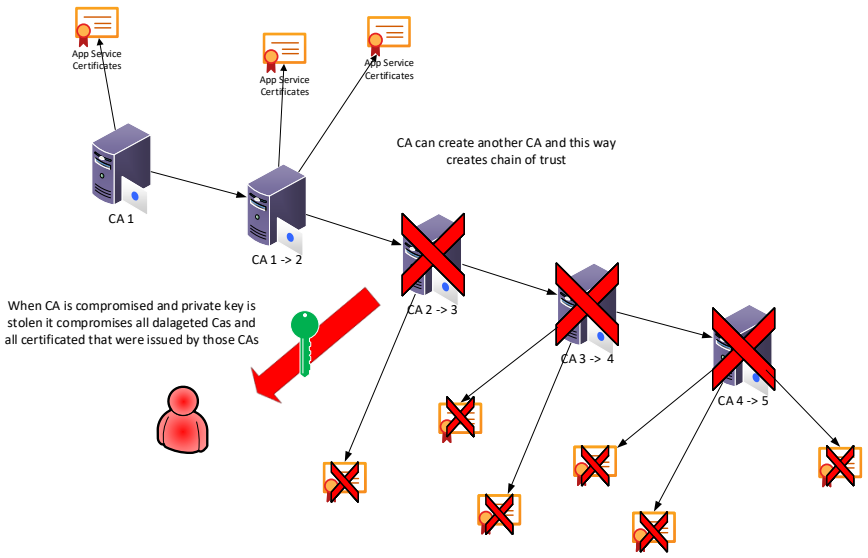

Figure 2.3: Chain of trust and compromised member of chain

2.3 Authentication

Authentication is process of proving user's identity. It is essential for maintaining security of any service. Authentication can be divided into 4 levels based on level of credibility of prove of user's identity. Those levels are called authentication assurance level (AAL) and are numbered from 1 to 4 when number 1 is indicating the least credible prove of identity and 4 the highest. Those levels are tied to sensitivity of systems where systems with low sensitivity requires low level of authentication assurance and so AAL 1 is recommended for them, for systems with high sensitivity or critical systems AAL 3 and 4 is recommended.

Identity of user can be proven via three different ways.

1. **What user knows** This way of authentication verifies knowledge of a secret shared between a service and user who is trying to access it. This

secrete must not be known to anyone else otherwise a whole process of authentication is compromised.

2. **What user has** This authentication method requires user to posses a security token issued by an owner of a service or by an other entity trusted by the owner of the service such as system administrator. A security token can have various forms e.g. a smart card or a flash drive or a look-up table which user receives during registration process. These days it is very popular to use cell phones as security token as an out-of-band authenticator. Token can contain either digital certificate signed by trusted CA (cards or flash drives) or can be cell phone which receives verification codes (SMS or via secured authentication apps) through independent channel from authenticating service or from other trusted service which is used in authentication mechanism.

3. **Who user is** User is authenticated based on his bio metrics e.g. fingerprints or corneal scan. This method is the safest one, however it is expensive and its implementation into environment is very demanding and it is not suitable for all situation where authentication is required.

These different ways of authentication can be combined together and used as multi factor authentication (MFA) which is the recommended standard these days for services which requires authentication assurance level (AAL) 2 or higher. Services requiring AAL 2 or higher are services with moderate or high sensitivity. [6]

2.3.1 Kerberos

Kerberos is an authentication protocol providing single sing on (SSO) functionality. Kerberos gathers network services and provides safe authentication mechanisms for them. All services imports their password databases to dedicated akey distribution center (KDC) and users are authenticating only towards the KDC server.

All services manged by the KDC and all additional KDCs which authenticate same services creates together a realm also one KDC can manage more than one domain.

Kerberos protocol eliminates necessity of sending a full text password or its hash while authenticating users to network services therefore it mitigates a possibility of eavesdropping a password or a password hash which can be than used in pass the hash attacks. Authentication mechanism using Kerberos protocol works in simple way. User during login to his local station or by using *kinit* command requests so called ticket-getting ticket (TGT) by providing his principal, which is user's unique identifier, with realm identifier and timestamp to a KDC which manages specified realm. If provided principal is correct KDC sands back to user 2 encrypted packets first containing session key, time to live

(TTL) of TGT etc.and its decryption key is hash of users password or users private key and TGT. The process of obtaining TGT is shown at 2.4. The decryption key for TGT has only the KDC which issued the ticket. TGTticket has expiration time in matter of hours at most lower tens of hours.

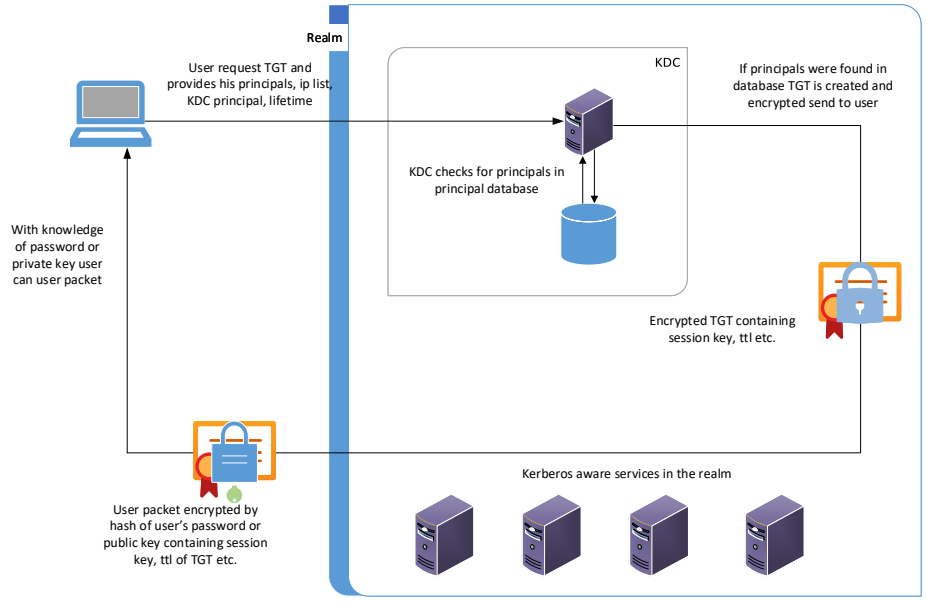

Figure 2.4: Process of obtaining TGT

The user packet and TGT contains these properties:

1. User packet
    - principal of KDC
    - timestamp
    - TTL of TGT
    - session key
2. TGT
    - client principal
    - KDC principal
    - IP list of hosts
    - timestamp

- TTL of TGT
- session key

Certificate authentication is not build in capability but can be added as an extension to Kerberos protocol.

For authentication to network services grouped in realm. User sends TGS requests to the KDC providing service domain name, his principal and timestamp and encryptes it all his by session key. In addition user attaches his TGT to the request. The KDC checks if user has correct session key, by decrypting request with session key from provided TGT, and provided service principal. If both are valid KDC sends back to user packet encrypted by session key and TGS encrypted by secrete key which is shared only between KDC and particular service. Obtained response has similar properties as response for TGT request main difference is that user packet from TGS response and TGS contains session key for communication between user and the service. In order to access the service user sends his principal and timestamp encrypted by service session key together with TGS as shown on 2.5

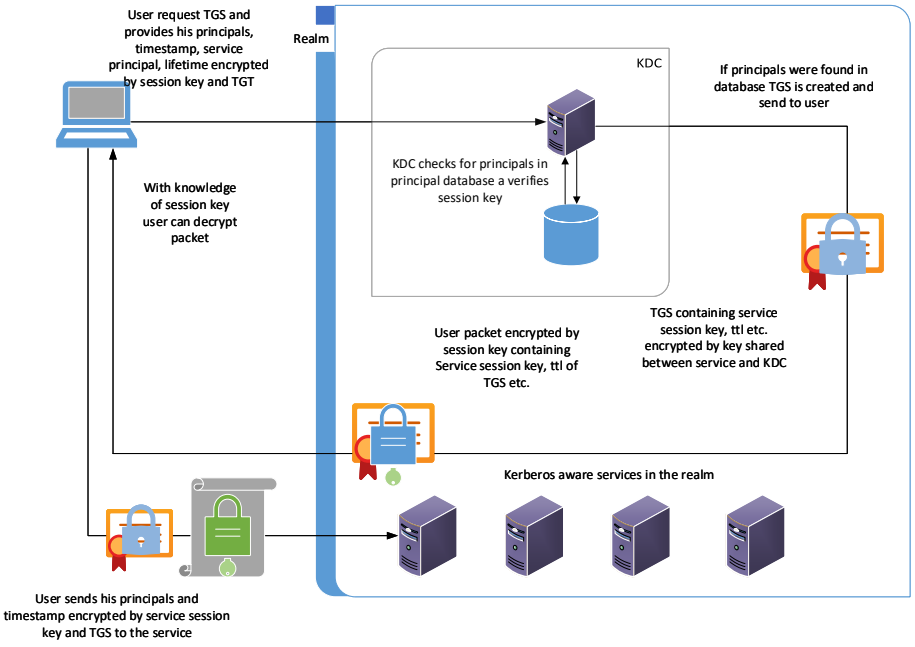

Figure 2.5: TGS request and authentication to a service

For Kerberos protocol to work properly it is necessary to provide domain name resolving service, this implies that at least */etc/hosts* file must be configured. However recommended practice is to deploy domain name service (DNS)

server. Domain name resolving must be set up because a ticket-granting service (TGS) request towards KDC contains domain name alias of a service as service principal. When TGS is being created then provided domain name alias is always resolved to an corresponding IP address then host name from address record is used into TGS.[7]

2.4 Access management

Authorization is a way of restricting or allocating resources of a system. These resources are allocated to individual users or groups of users based on attributes tied to their identity or to particular resources.

Access management can be divided into two parts:

- access control
- privilege management

Both of these parts are dependent on each other.

2.4.1 Access control and Privilege management

Privilege management manages attributes such as role of account, group membership or any other data which are necessary for particular system or individual entities as well as provides and manages policies which authorizes given entity based on its attributes.

Access control defines which attributes are used for user authorisation. Access control can be further dived into those three categories which are explained below this list.[8]

1. Identity-based access control (IBAC)
2. Role-based access control (RBAC)
3. Attribute-based access control (ABAC)

2.4.1.1 Identity-based access control

IBAC is the most basic access control method where main attribute for dedicating resources is user's identity. Problem of this method lies in its maintenance when deployed in larger scale. Assigning rights based on identity means maintaining each account individually.[8]

2.4.1.2 Role-based access control

RBAC uses roles for granting privileges. Those roles are created based on possible task which can be done on a system. Based on those use cases and according to least privilege paradigm privileges are assigned to particular roles. Then when account is created a role or roles are assigned to it based on its purpose.[8]

This approach has significant advantage against IBAC method because it requires fewer entities which has to be maintained when managing access privileges.

RBAC compared to IBAC adds one more step to system deployment process. Before a system with RBAC is deployed all possible use cases of user activity on the system should be considered, then appropriate roles must be created. This step can be really demanding but its crucial for proper functioning of RBAC model and security of the system.[8]

2.4.1.3 Attribute-based access control

ABAC is the most complex way of access control. Privileges are granted based on attributes provided by user and by system providing its resources as well. Those attributes are matched against policies designed to evaluate those attributes and based on results privileges are assigned to user.

However complexity and sensitivity of ABAC comes with its prize in form of administration which grows according to number of attributes and policies which needs to be maintained. Second problem is that not all system are supporting this kind access control these days.

As improvement of ABAC risk-adaptable access control (RAdAC) was proposed. It adds current conditions into consideration however RAdAC-based system must have human decision making incorporated in their policies in case of unexpected situations. Due to variety of scenarios policies must be designed with extreme caution not to block all users from accessing a system or similar cases.[8]

2.4.2 Privileged access management (PAM)

User accounts can be basically divided into two groups high risk accounts and low risk accounts. Low risk accounts are every day use accounts or some service accounts with low privileges. Those accounts should have standard protection such as strong passwords, blocked security configuration options etc.

High risk accounts or privileged accounts are accounts with access to some sensitive information or configuration privileges either local, domain wide or server management access. Those accounts should be closely monitored and well protected. Comfort of their use is not the main concern but it should be thought of. Use o those accounts should be allowed only from dedicated

secured hosts called privileged access workstation (PAW), to eliminate leakage of credentials or other sensitive data by using less controlled and less secured host and for better protection of accessed systems in form of more strict firewall rules. Those accounts should have clearly defined life cycle and should be tied to one particular user only. All aspects of use of these accounts should be clearly defined such password strength which should be greater than minimal recommendation which are 8 characters length with check for breached and known passwords (random generated at least 16 characters), password reset policy, creation, deletion of an account and credentials handover. On those accounts least privileged paradigm is applied as well. Accounts dedicated to particular tasks should have privileges only sufficient for those tasks.

With those accounts well protected it is good to think about aspect of user comfort. A good example can be logging in to those account using and extremely hard long passwords. This can cause problems in more areas. It can trigger false positive warnings in monitoring systems which track activity of those accounts when those passwords are misstyped. It forces users to write down passwords in order to remember them or they could be recorded by key logger when frequently inserted through keyboard. Therefore it is good practice to employ some password management mechanism which allows auto-type and storing of passwords. Some more advanced solutions can even separate user from a password to level that user can not view the password and keeps it a secret, and only logs-in previously authenticated user under desired account.

Privileged accounts should be very closely monitored therefore it is recommended to deploy some kind of central log management such as Elastic Search which is deployed now in ATLAS or Splunk or even better some security information and event management (SIEM) solution such as IBM QRadar. SIEM systems are capable of more advanced correlations above incoming log records and can be integrated with other security software such as vulnerability scanners, other monitoring systems and thread databases and correlate all gathered information together in additions SIEM systems are equipped with great number of security correlation rules straight out of the box.

2.5 Data rights management (DRM)

Goal of DRM is to protect data from unauthorized sharing and making unauthorized copies. There are many products trying to achieve this goal through various technologies specialised for different types of files such as Microsoft Information Rights management which protects office files and emails when Microsoft Exchange server with Active Directory (AD) is deployed or Google Widevine which specialises on multimedia content and many more.

However, all solutions have somethings in common they all use some form of encryption in combination with connection to servers to retrieve a decryp-

tion key but, everybody implements it differently. Some solutions need to have installed clients applications, some need connection to key servers every time protected content is accessed this is called always-online DRM, others have option to temporarily save decryption keys and are able to operate without connection to key servers for some time, some implementations uses metadata attached to protected files others incorporates their data into protected files themselves etc. All implementations are designed for particular formats only and not for general data or custom formats.[9]

2.6 Remote access

Remote access can be defined as possibility to telecommute with a system or program from places elsewhere than organization premises. There are multiple techniques of remote access. Each method has its pros and cons. They all have few things in common.

- They are secure as long as end device is not infected or otherwise tampered with.
- Offer various authentication mechanisms allowing them to use current authentication scheme or create a new one.
- Encrypt communication between end point device accessed system or network (Some protocols does not use encrypted communication but its highly recommended those which does).
- Provides way of copying data from remote station to local device (this option can be disabled in justified cases).

Remote access methods can be distinguished based on high-level design and can be divided into tunneling such as virtual private network (VPN)s and secure shell (SSH), remote desktop access, portals and direct application access.[10]

2.6.1 Tunneling

Tunneling is technique when encrypted tunnel through public network is created between user's end device and accessed system and all communication goes through that tunnel.

VPNs can be distinguished based on several criteria such as purpose, encryption technology etc. For purpose of this thesis VPNs will be divided base of their use to site-to-site VPN which is used to securely interconnect two separated networks to behave as one and remote access VPN which is used for accessing remote networks from individual end devices.

Difference between remote access VPN and SSH is that when local end device establishes connection with remote access VPN gateway tunnel is created

user gains access to remote network where user must access desired system and all his communication goes through established tunnel. When user is connected through a SSH tunnel, in general, connection is made to particular host which is running SSH daemon and user interacts with that particular system using shell interpreter and depends on user authorisation which actions can be done by him. In case of VPN, user does not interact with VPN gateway except from authentication, SSH can be also used as a encrypted tunnel for other applications when network ports of those applications are redirected to SSH port (usually port 22) and same on the remote ssh server. This implicates that SSH server can also be setup as gateway but all communication which is supposed to go through ssh tunnel must be manually redirected to the tunnel. In general, communication between any gateway and internal network is not protected if no other precautions are taken.[10]

2.6.2 Remote desktop access

Remote desktop access gives the user the same options as if he would sit in front of a real device except physical access to hardware. Communication between remote device and local endpoint is encrypted. This method has advantage for common user as it provides graphical user interface (GUI) if the system provides one. Compare to SSH which providing only shell interpreter. However, in case when remote station does not have GUI it offers similar possibilities as SSH. Remote desktop access provides little bit more possibilities and restrictions which are applicable on user because local policies on stations are applied on remote user as well. As example of advantage of remote desktop access is possibility to use clip-board.[10]

2.6.3 Application portals and direct application access

Application portals are servers providing centralized interface for interaction with one or more applications. This interface can be web-based or can be in form of installed client app on local device such as terminal server client. Server running application portal also runs application clients which communicates with application servers.

In case of direct application access user access remote application directly through web-based application portal or through application client installed on local station.

Securing connection and user authentication between local host and an application portal or an application must be implemented in an application or portal itself. This solution requires application servers to be accessible from the internet and so to be placed on a network perimeter. This requirement means that only low risk application should use this method of remote access.[10]

2.7 Jump server

Jump server is remote access server which works as intermediate step when accessing a system. Jump servers have great application in security area. They can be used to protect privileged accounts or to disable use of disallowed software in an environment. It can be also use as a single point of access, which has significant advantage when investigating an audit trail after incident.[11]

2.8 Bastion host

Bastion host is host which is specially designed to withstand cyber-attacks and if breached caused minimal damage to an environment. It is used to protect sensible systems which needs to be remotely accessed. Role of bastion host can have many hosts in an environment such as proxy servers, web servers or VPN gateways.[12]

Chapter 3

Introduction to infrastructure

3.1 Worldwide LHC Computing Grid

CERN was founded in 1954 and since the beginning it brought together scientists from all around Europe. Since first experiment, which started in 1957 there was need to share measured data and so in 1989 first draft for World Wide Web was submitted then by the end of 1990 first web server was deployed.

By the end of 1990 when all four experiments on LHC were under construction it was realised that needed computing resources for data processing and simulations and data storage capacities were far beyond what could be founded by one organisation. Fortunately in late 90's the Monarch model was published introducing model of data centers divided into centralised tires with one central node Tire-0 and big regional data centers as Tire 1 and smaller data centers supporting individual Tire1 data centers in Tire 2. At the approximately same time I.Foster and C. Kesselman came up with idea of distributed computing model called The Grid which. It first introduced an idea of distributing workload between several computing centers however acting like one big computer center for an end user. Individual participants in The Grid such as users, computing centers would be connected and orchestrated by layer of software called middleware.

In year 2000 projects for exploring The Grid possibilities and developing middleware started e.g. European Data Grid (EDG) and later Enabling Grids for E-science (EGEE) founded by European Union (EU) and Open Science Grid (OSG) in the US.

In year 2002 the LHC Computing Grid (LCG) was found in order to coordinate development among EDG, the experiments hosted in CERN, and be able to incorporate other computing centers into project. LCG has hierarchical structure where participating smaller grids or computing centers coordinate their own resources and LCG orchestrates participating parties. Later collaboration with European NorduGrid and American OSG was started to

emphasise this worldwide collaboration the LCG was renamed to Worldwide LHC Computing Grid (WLCG).

With individual experiments located in CERN and computer infrastructure which is handling filtered raw located there as well, CERN computing farms became naturally Tire-0 of emerging computing grid. Tire-0 stores raw data and provides them to Tire-1 computing centers which holds backup of raw data. Tire-1 data centers holds all data which resulted from further raw data processing and from Monte Carlo (MC) simulations. Tire-2 data centers hold backup of all other than raw data and hold data requested by users.

Majority of computer power of all tires is used for running simulations. Workload distribution between sites is based on pull model. It means there is central queue which holds workload and sends pilot jobs to individual work nodes. Pilot job is simple script which finds out computing and memory resources of the node available and matches them against requirements of real job.[13]

3.1.1 Virtual Organisations

Virtual Organization (VO) is an administrative tool to distinguish researchers from individual experiments. From side of a grid VO is an administration unit to which resources are allocated. VO can be further divided into groups, those are managed by VO administrators. Resources can be further distributed into groups or individuals and other restrictions or privileges can be applied on them by VO administrator.[14]

3.2 ATLAS experiment

The Experiment must process enormous amount of data. The system which enables to process all recorded data is called Trigger system. Trigger system is divided into two parts called Layer 1 Trigger and High Level Trigger. As all the other systems on ATLAS detector they are developed for specific ATLAS needs. Process when ATLAS detector collects data is called online data taking and during that time it is crucial that all participating systems works flawlessly. For purpose of development and for solving emergency situations when error occurs those systems must be remotely accessible to minimize possible loss of data.

Online data taking process is watched by group of shifters from Control Room. Each shifter is specialised on particular component of the experiment or on particular process. Even though shifters are specialised on particular component of the detector they do not have to be experts. Experts or developers of particular components can be situated anywhere around the world, there comes the need of remote access.

Main job of shifters is to watch for system errors and miss-behaving of the systems and perform system specific tasks such as enabling and disabling

various sub-detectors during a run. This is done by shifter who takes care of Trigger system by inserting prepared configurations into Trigger during run as conditions change.

Set up of the Trigger system is done through 3 keys which defines systems behavior.

1. Supermaster key which defines possible settings of trigger system.
2. L1 Prescale Set key which set sets up L1 Trigger.
3. HLT Prescale Set key.

Keys 2 and 3 can be changed during a run based on expected change in luminosity or in case of unexpected conditions. Key sets are prepared by working groups based on simulations and expected values during data taking. In case of unexpected conditions shifter inserts different than prepared keys from the menu based on consultation with Trigger expert and Shift Leader.

There is one more additional key, which defines LHC fills parameters (parameters of accelerated particles), this key is called Bunch Group Key and can be change during a run as well.

Whole shift is under command of Shift Leader who is responsible for whole process of data taking and for assessing irregular request like remote or physical access to the online data-taking systems.[15, 16, 17]

3.2.1 L1 Trigger

L1 Trigger is costume made hardware dedicated to select incoming signals from ATLAS sub-detectors for further processing. Analog signals are first converted to digital, which are then processed by L1 Trigger. It is set up by previously mentioned L1 Prescale Set key which defines which sub-detectors should be active and based on what criteria incoming data should be processed. L1 Trigger reduces the frequency of incoming data from 40 MHz to approximately 100 kHz.[18]

3.2.2 High Level Trigger

The High Level Trigger, also called the Software Trigger, is grate computer farm placed in cavern of ATLAS experiment. Its main purpose is to make fast reconstruction of incoming events and with further analysis decides which events are going to be stored and which not. High Level Trigger sorts data based on approximately 1500 defined selection chains. This farm is using software developed by ATLAS collaboration to reconstruct whole events or just the particular areas of interest. One reconstruction can not take more than 500 ms. High Level Trigger reduces rate of incoming data from 100 kHz to 1 kHz in average which is than stored. To cover possible peaks in rate of incoming data, data are first written into buffers and then into storage.[18]

3.3 Policies overview

This section provides brief insight into some business process which might have impact on security in ATLAS environment.

3.3.1 Publications based on ATLAS data

All publication based on data taken by ATLAS experiment or from MC simulations or regarding ATLAS infrastructure must undergo ATLAS Collaboration approval process.

The approval of publication is complex process. Before submitting a paper an Editorial committee must be established. Editorial committee includes members of ATLAS Collaboration from different areas of interest, members of team working on a publication, and member or formal member of Publication committee. When first draft of a paper is ready it is posted on CERN document server for commenting by other ATLAS members. The publication must already include all data, plots, and tables in the publication. Comments with significant impact must be answered and reasoned by member of author team. This process is called circulation. After all comments are processed, either incorporated to the paper or reasoned why they were disclaimed the publication undergoes second round of circulation. After final changes are made based on comments from second circulation and approval from Publication committee chair and consultation with Editorial committee, the paper is moved for approval to ATLAS Spokesperson and CERN management. In case of comments which would suggest significant changes to publication another circulation might be suggested.

After the publication is approved in all ATLAS bodies and by CERN management, it is send to chosen journal. All notes from journal are send to contact editor (member of team of authors) and to Editorial committee for consideration. If incorporating of those comments would make significant changes the approval process can be invoked again.[19]

3.3.2 Gaining access to WLCG

To be able to access the WLCG, user must first obtain X509 personal digital certificate and be part of recognised VO.

To obtain certificate, user must first request certificate from his national CA or from CERN CA. Before certificate is issued by national CA user must prove his identity, using passport or other valid form of ID, to registration authority in case of Czech Republic CA and registration authority is CESNET organisation. Certificates issued by national CA are valid for one year then they must be renewed. Personal certificates signed by CERN CA are available only to users with certain types of CERN accounts.[20]

Next step is to become member of WLCG recognised VO, in this case ATLAS VO. To become member user must first have CERN account.

To register for CERN account, user must fill the form which signed by team leader or deputy team leader and with copy of passport is send to ATLAS secretariat. After account is registered CERN credentials are send to a requester using email filled in the form.[21]

User then registers to ATLAS VO on WLCG web page inserting his personal certificate user name and email which is associated with his CERN account.[20]

3.3.3 Remote access to sensitive systems

Currently deployed solution of remote access uses SSH. An user connects directly to desired host. When connecting the user is authenticated against ATLAS Kerberos key distribution center using his password.

Personal accounts are used for every day use as well as for administration of systems and user accounts. There are no dedicated accounts administration accounts or accounts dedicated to other task requiring elevated rights.

When accessing online data taking systems during data taking, this access must be approved by shift leader sitting in the ATLAS control room. Shift leader accepts or denies request based on identity of user and based on content of message which is attached to a request. The request is send as part of authentication process on online data taking systems.[22]

Chapter 4

Analysis of current situation in information security

This section provides brief assessment of ATLAS infrastructure and processes from security point of view. Its goal is to pin point potential risks which should be minimised by design proposed in chapter 5.

Overall security precautions taken by ATLAS collaboration works well but there are some weak spots.

The greatest weakness, which can not be mitigated, is that there is no way to control what kind of stations is used by users. The situation does not allow to enforce any security standards and precautions from side of ATLAS. Accessing stations do not have any central monitoring or thread detection systems to protect them or detect malicious software and so they should be considered untrustworthy from side of ATLAS and their access to environment should be minimized.

Another danger of not controlled end hosts is that no security policies can be enforced such as minimal passwords requirements or other basic security rules.

In ideal situation ATLAS Collaboration should lend personal station to all users who are accessing ATLAS environment to be able to control security standards and installed software and restrict users privileges on those stations.

4.1 Remote access assessment

Remote access through SSH is, in general, secure way of communication however SSH is only communication tool. Whole process of remote access and management of remote systems can be compromised from other directions.

First security issue, which can end up in compromising whole ATLAS environment but it also provides opportunity for performing simple denial of service (DOS) attack, is the possibility to access sensitive systems directly

from the internet. In case of an attack this can lead to lost of valuable data from online data taking or other bad scenarios.

Situation when host with enabled remote access must also have public IP address is not secure and usually opens more than one possible attack vector. At first, with all the systems on network perimeter all of those systems act as bastion hosts and so there is much greater chance of some system being compromised. Current situation puts greater pressure on maintenance and security of a system. Also, whole ATLAS environment loses one additional layer of security which is provided by closed network environment. This situation makes development and upgrades of those systems more demanding and increases requirements on authentication mechanisms and quality of access credentials and all deployed software. In addition, it puts greater load on firewalls protecting those system in form of more complex sets of firewall rules. In case those firewalls are running on local hosts and not on dedicated host, it can consume great amount of computing resources needed elsewhere.

4.1.1 Authentication to remote hosts

Kerberos as authentication protocol is currently one of best possible solutions available. It offers SSO which reduces threat of password exposure when authenticating to individual services. It also mitigates necessity transferring password or its hash over a network.

The weakest point of whole authentication mechanism is use of password in combination with no other factor. In general passwords are not the best authentication methods because users must remember. This causes that passwords can not be too long nor too difficult. This also causes there is always some pattern in their structure. All those things significantly lower number of possible combinations which could be the password and make easier to guess it. There for second authentication factor should be used which is independent on the first one. This ensures that even when password is stolen or guessed an account is not breached.

With addition of second factor to authentication scheme a password still remains the weakest spot of authentication. To improve even more authentication mechanism, password should be exchanged for more secure form of authentication such as certificate or bio-metrics. Certificates have advantage in relieving user from creating password which is necessary to remember and so it is relatively easy to guess. Creating difficult passwords can end up in scenarios like one sufficient password is used over and over again creating risk when broken great amount of accounts is compromised or other worse scenarios is when passwords are written on papers etc.

4.1.2 Privileged account use

Not separating privileged and every day use accounts is very comfortable and user friendly. However it denies completely security in depth paradigm because when this kind of account is breached it can jeopardise whole environment.

Every day use accounts are exposed to various threats. It is caused by nature of their use therefore it is likely that they will be compromised and so the goal should be to minimize damage when this happens. Another reason why to separate every day use accounts and privileged accounts is because of their monitoring. Every day use account creates thousands of log entries. This means that overall amount of log entries is much greater than number of log entries from account which is used just for small fraction of operations. This makes much easier to follow audit trail of dedicated privileged account than trying to extract audit trail from thousands of unrelated log entries.

Authorising accounts for access to data taking systems by shift leader during LHC run does not seem as a good way of elevating privileges of requesting accounts. It relies on knowledge of a particular system by Shift Leader and trust in user who is attempting to access the system. This trust should be given only to users who know system which they are accessing well enough not to compromise its functioning. Other users who need access to those system should be allowed access only when those systems are not in action. This does not have to be decided by Shift Leader on the fly but it can be assessed by some kind of committee and after careful consideration user could be granted privileged account with all time access to particular system. This would imply creation of two groups of privileged accounts, one with all time access and one with access only when the systems are in standby mode.

4.2 Protection of data on WLCG

The weakest spot in protection of recorded data lies in possibility to download them. This problem is very difficult to overcome and every possible solution which keeps this possibility will have potential weak areas. In order to minimize those weak areas all possible precautions should be taken.

4.2.1 WLCG authentication

First of all, strong authentication mechanism should be used. Right know authentication is done by using personal certificate which is very secure in order to protect accounts from being compromised by brute-force attack or some other type of attack aimed to guess account credentials. However, it does not protect the account in case when host with installed certificate has been compromised by malware or stolen. In this case an account is not secured at all. As mentioned before, ATLAS staff uses personal stations where no

security policies can be enforced so they all should be considered as potentially compromised.

4.2.2 Downloaded data protection

Once data are downloaded from the WLCG, ATLAS instantly looses control over that copy of data. To bypass this problem it is necessary to alter downloaded data to a form in which it would be unreadable without ATLAS cooperation. In this case it implies to encrypt data and the decryption key should be held by ATLAS Collaboration only. This brings new challenges in form of how to safely lend the decryption key to user who is authorised access the data. For how long to lend this key and in case it should be for longer time, how to store it on personal stations with the smallest possible chance for the key being exposed.

Chapter 5

Suggested solution based on current situation

This section is describing design of possible improvements in information security of chosen systems and in addition brings higher level of automation to Trigger System remote access. All suggested solutions take in consideration already used technologies. Some of the suggestions could be and should be applied to whole environment except systems and users where it is not possible for objective reasons. Those systems should be known and monitored even closely than rest of the environment because they are potential vulnerabilities in the ATLAS environment.

5.1 Remote access to online data taking systems

In order to ensure availability of online data taking system all the time, whole solution must be deployed in high availability (HA) mode. In order to use this solution more effectively high availability should be implemented by using two tcp proxy servers with virtual IP address and primary and secondary component as shown on picture 5.1.2 with VPN gateway.

As mentioned in section 4.1.2, remote access to online data taking systems should be divided into two categories one with all time access and second with access granted only when systems are in standby mode. This can be tricky to implement because RBAC model can not assess current state of a systems. However, this problem can be overcome by managing access to credentials of privileged account. This can be achieved by creating credential database administered by ATLAS mentioned in section 5.1.2 later. When online data taking systems start, it triggers script which logs out all users with accounts from second group and hides a table or file (depends on type of database)in database which contains credentials for those accounts.

5.1.1 Single point of entry

In order to improve situation when ATLAS systems are on network perimeter, the only one point of entry into the ATLAS environment should be created. All systems would be accessible through this entry point except those systems are meant to be accessible from the internet, such as email server or web servers. However, only necessary ports for proper functioning should be opened such as 443 for HTTPS for web server. All entry points into the ATLAS network should be monitored by thread detection system and protected by firewall and all activity.

An entry point into network for users should be bastion host which is either VPN gateway, SSH jump server or remote desktop access jump server which helps establish secure connection between local host and ATLAS environment. All this solutions allow firewall rules in the ATLAS environment to be much simpler and takes off CPU load which is taken by processing incoming network traffic. This entry point into network is dedicated to protect access into network, and therefore it can be set up with maximum focus on security.

5.1.2 Use of jump server

A jump server should be implemented with each earlier mentioned remote access technology to be used with privileged accounts. A jump server should be separated from a bastion host because it has access to encrypted database with credentials to privileged accounts. This database can be stored on dedicated credential server but in order to save resources and considering that this jump server is not be accessible from the internet, it is acceptable risk to store the credential database on network storage with at least RAID 5. This storage is accessible from both instances of jump server. This solution allows great variability when managing privileged accounts. For example, this kind of deployment allows to implement login to privileged accounts without knowing the password. This kind of solution minimises the risk of credential exposure.This effect can be achieved by deploying CyberArk solution in the environment. CyberArk is commercial tool for privilege access management.

There is few other reasons why it is good idea to use jump server for access to privilege accounts.

None of the ATLAS staff is provided with personal stations which would be govern by ATLAS, therefore there is no control over local stations which are accessing ATLAS network. There are no security standards and policies which could be enforced on those personal stations, no way to ensure that all security patches are applied and no way to monitor that personal station is not infected with malware such as worm or keylogger. In order to protect credentials of privileged accounts, sensitive systems should be access through jump server administered by ATLAS collaboration.

Second very important reason for deploying a jump server is the possibility

to use it as PAW. PAW allows to create policies which denies all logins to sensitive systems from other host than PAW. Users accessing PAW in order to access sensitive systems under privileged accounts have minimal privileges on PAW.

This system of picking up credentials for privileged accounts has other advantages such as it adds one more layer of protection when using privileged account in form of authentication to the jump server another big advantage is that passwords for privilege accounts can be changed without cooperation with owners of accounts. If user needs to use his privileged account, he just logins to jump server picks up his credentials and from jump server logs in to desired systems. This set up allows change of passwords to privilege accounts on regular basis.

Third reason, why to use PAW, is monitoring of privileged accounts. With single point of origin it is much easier to monitor use of privilege accounts and so discover their abuse or attack on an account either successful or not. To be able to appreciate this advantage, there must be some monitoring system deployed and properly set up which will be talked in section 5.1.3.

5.1.3 Monitoring

One of the goals of this thesis is to automate remote access to online data taking systems. This implies that some accounts will have access to those systems at any time for maintenance or in case of error or any other justified reason. Therefore, it is very important to monitor those accounts same as all other accounts with elevated privileges. For this purpose a SIEM solution is used. This system would collect log messages from jump server online data taking systems and bastion host which should provide entry point into network and could collect log messages from any other system which would be necessary to watch.

SIEM system saves all the incoming information and processes them according to defined rules. Those rules can monitor accounts activity on various systems correlate it together. It is able to follow patterns in accounts activity and system behaviour and can report deviations from standard and many more thinks.

Properly set up SIEM system works as automated monitoring of user and system activity which enables detecting of security breach, malicious behaviour. Except detection mechanism, it works as tool for investigating of reported suspicious activity.

Difference between SIEM system and log mangement system, such as Elastic search which is deployed now, is that SIEM can correlate on broader scope of data and from the box contains set of hundreds of predefined rules which are focused on securing an environment. It also provides long term user behavior analysis and evaluation. Another advantage is the possibility to connect SIEM

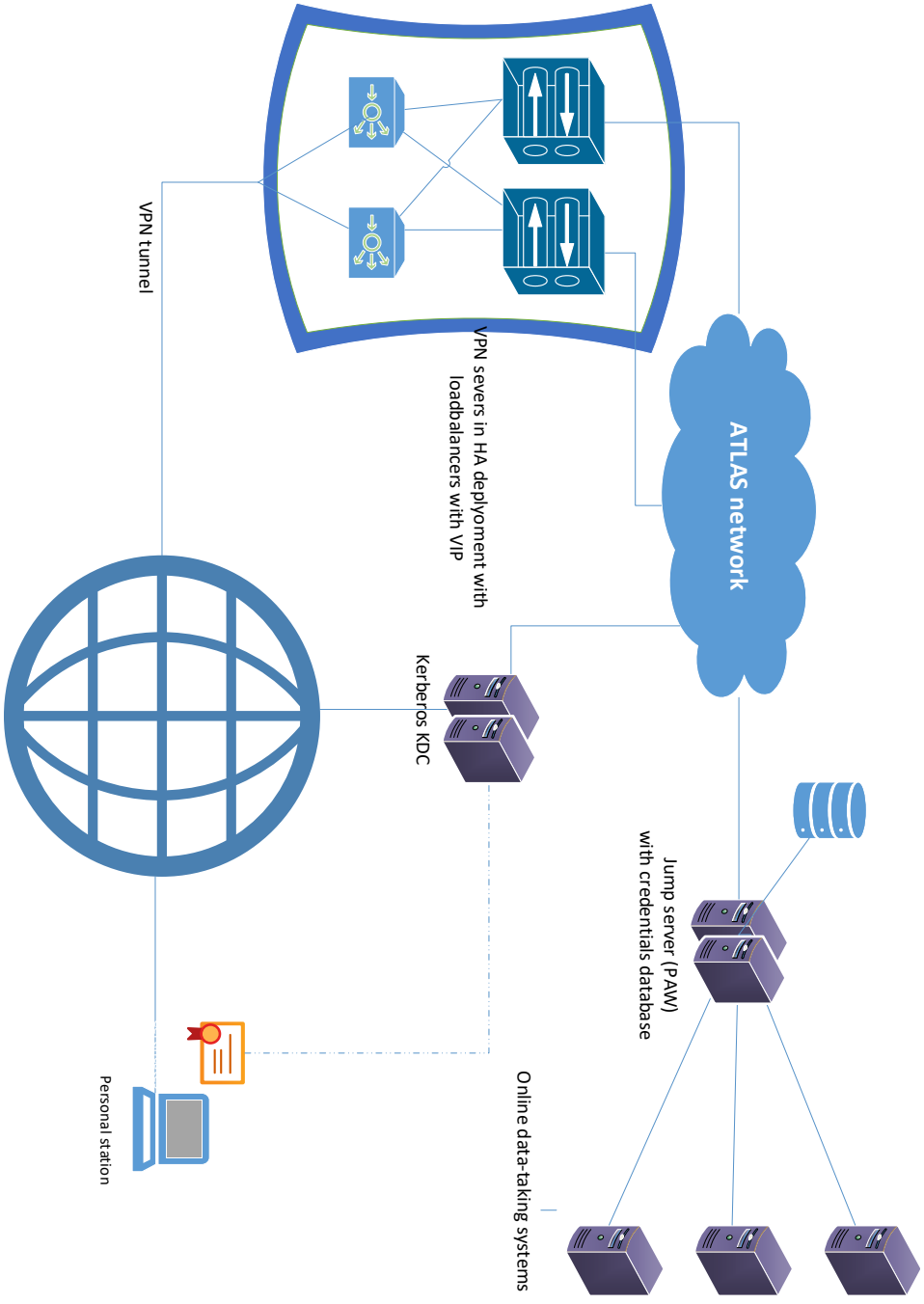

Figure 5.1: Remote access design in HA solution with credentials database.

solution to thread database which feeds it with up to date information about malicious web addresses known signs of malware etc.

5.1.4 Authentication mechanisms

In order improve authentication to be more secure, there are two things which can be done.

The first thing is to deploy two factor authentication. It has became standard these days. This can be done by several combinations of authentication mechanisms. However, there is already use of trustworthy certificates to authenticate to WLCG and so those certificates could be used to authenticate to ATLAS internal network as well. This mitigates the use of passwords which is previously mentioned as the weakest point. As the second factor Hardware security module (HSM) token like a smart card or secured flash drive which would carry the certificate should be used. In addition, this solution would protect the certificate which is now installed on personal computers of participating researches where is no control over security set up.

ATLAS collaboration would have to administer those HSM tokens. In order to simplify the administration of those tokens, ATLAS should became the CA which signs certificates used for access to WLCG and to ATLAS environment instead of national CA.

5.1.5 Privileged account lifecycle and use

Use of privileged accounts should be restricted only to intended purposes. For every day use dedicated accounts should be used with least privileges possible. This is enforced by monitoring of those accounts in SIEM system and by access policies.

Account management is important part of a system administration, it becomes even more important when administrating privileged accounts. To make administration of those accounts more efficient RBAC or ABAC method of access control should be deployed. Even though ABAC method would offer more precise and dynamically delegated access rights, use of ABAC would implicate to overcome several major problem. This makes use of this method almost unusable.

First major problem is that there is no support in general for this method. That means this access control method would have to be newly developed on majority of systems. The second obstacle of this approach is that this approach needs man assistance in order not to unintentionally lock out all users from a system.

The RBAC method is less precise than ABAC but RBAC is used on majority of systems and is fully automated but it requires more aware users because privileged accounts have the same privileges without consideration of a status of a system.

However some advanced control of access control must be deployed than just IBAC because there is too many accounts that must be managed. Number of accounts is not just given by number of administrators. Every privileged account must be tied to just one user and one or small group of tasks that usually means multiple privileged accounts per administrator. This precaution is important for audit trail which is left after each action done by privileged account to be unequivocal. This means there will be many privileged accounts to manage.

Creation of privileged accounts should have clearly defined procedure which must be strictly followed. This procedure should be robust enough to uncover all possible unjustified request.

The procedure could look like this.

1. User log is to ticketing application and creates ticket requesting privileged account with reasons why this account should be created.

2. The ticket is forwarded to user's superior and to ATLAS dedicated security ATLAS admin. Who both must approve creation. of this account. It is expected that for both people this request will be expected.

3. Approved ticket is provided to ATLAS administrator who creates requested account.

4. Newly created credentials are then loaded to a databased on jump server.

5. In case user did not have privileged account before new account on jump server is created. This account uses SSO authentication.

6. Password to database with credentials should be send to requester via separate channel like text message. This password should be change after first log in.

Creation of a new privileged account requires to have privileged account as well.

Clearly defined policies should be created for deleting a privileged account and all other administrative actions with privileged accounts. Process of deleting privileged account does not requires approval from users superior and security administrator but it should involve verification of this action via different channel like mobile phone or email.

5.2 Data protection on Worldwide LHC Computing Grid

To protect taken and simulated data on the WLCG, there are two approaches to this task.

Data access would be well secured through authentication via personal certificates with addition of second factor. With addition to HSM tokens mentioned in previous section the authentication mechanism gains desired level of security. The issue here lies in impossibility to revoke access to already downloaded data from the WLCG in case when an individual or institution looses the privilege to access the data. However, all solutions requires at least periodical internet connection and all solution requires data to be encrypted and decryption key to be stored on ATLAS private servers.

Both approaches have the same basic idea and differ only on what kind of host would a client program installed on and slightly on authentication scheme which would be used. The two models would be.

- Server dedicated to work with downloaded data from the WLCG. Data from the WLCG would be downloadable only to this dedicated server. A participating institution in order to be able download data form the WLCG would have to deploy dedicated server for this purpose. Researchers from this institution would not be able to copy data from this server to any other host. In order to decrypt downloaded data the server would authenticated using installed trusted certificate. This server would have installed client application developed by ATLAS for managing requested decryption key. This decryption key would have only short time to live (TTL) approximately 10 minutes.

- Data are downloaded to personal computers of researchers. Individual researchers would have to install client application in order to decrypt data downloaded from WLCG. Authentication scheme would be the same as in case with the server. User would use his personal certificate on HMS token to authenticate. Downloaded key would have longer TTL than in case of server approximately 2 days then the user must request decryption key again. Longer TTL is in case users need to access downloaded data offline.

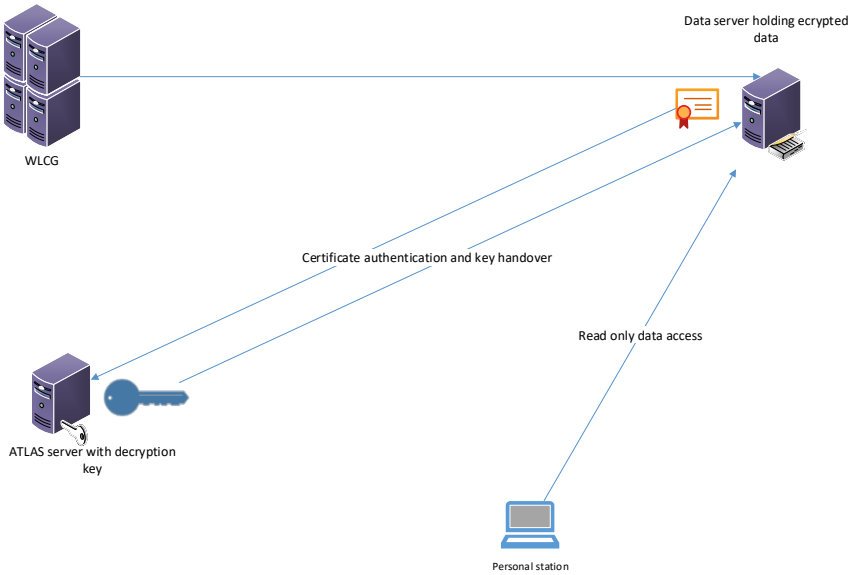

Figure 5.2: Access to downloaded data through a institution dedicated server.

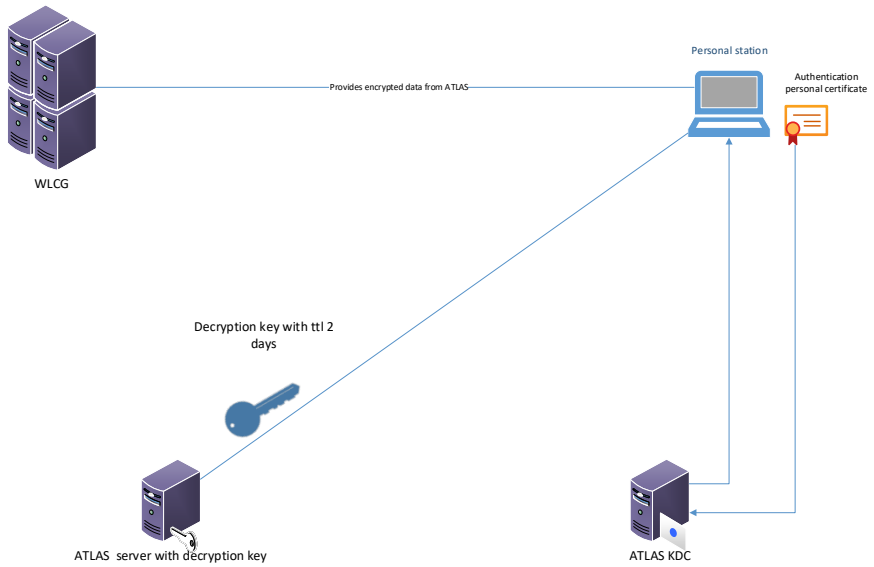

Figure 5.3: Access to downloaded data on personal station.

Both models need to handle data and specially decryption key in secure way. Therefore, after authentication to server providing a decryption key the key is send encrypted, using requester's public key contained in certificate provided by requester. In case of server solution the, key would be saved only temporally in RAM memory and was deleted when the client program ends or when TTL expires. In case of personal stations the decryption key would have to be to saved on local storage and so first it would be encrypted using security processor which is build in almost all modern computers and contains inextractable encryption key. This minimizes the possibility to extract the key from computers memory.

Data decryption should only be done on data which are currently requested not on all and decrypted data could not be saved on hard drive, it should be kept only in memory of program and deleted immediately after their use.

In case of both models, data revocation should be very easy. When authentication to server providing decryption is disabled in the worst case scenario, small portion of data is accessible for approximately 3 days, then the local decryption key expires and all downloaded data are inaccessible.

In case of publications involving data, a publication before approval should just reference necessary data so that reviewers could access data on the WLCG. After publication is approved then necessary data would be released by the ATLAS staff in decrypted form to the author, in order to be included in publication.

Conclusion

This thesis had 4 main goals. The first one was to assess the current situation of remote access to online data taking systems from the information security point of view. The second one was to assess the situation of access and management of data created by ATLAS experiment and 3rd and 4th goals were to suggest improvements which can be made on analyzed systems.

First step was to analyze both systems, this step revealed some areas where systems both systems could be improved. From those areas some deserve to be mentioned. The most important one, which involves both areas included in this thesis, is no possibility to enforce use of ATLAS managed endpoint stations by the end users. This impel to think about all user station as potentially risky and take that in consideration when creating a new businesses process.

On the remote access to online data taking systems process, the weakest spot are that critical systems have open ports into the internet. The second weak spot is in account management and lies in not separating privileged accounts and basic everyday use accounts. Both aspects pose thread to the whole environment. In following part high level design of possible solution was outlined. Solution consists of creating single point of entry into ATLAS environment and hiding all possible systems behind this single point of entry so those systems are not publicly accessible. As response to the second weak spot, the proposal was made to create central managed privileged accounts for purpose of administrating online data taking systems. There were some other security precautions suggested to further improve security of suggested improvements and the whole remote access process.

The data access and data management part showed one weak spot and that is the possibility to download data to local storage. Given by the nature of data, there is no commercial or open source product which would be able to apply any DRM technique and so this brings a new challenge of developing of a custom made solution for the ATLAS. This thesis offers possible technology independent approach of implementing one of DRM techniques into ATLAS

environment, which consist of necessity to retrieve a decryption key from the owner of a data, in order to access them. The process of implementing this will require detailed implementation plan.

The thesis was able to fully complete 3 out of 4 main goals. Analysis of both business process, remote access to online data taking systems and the data access and the data management and finding possible improvements for those business processes. In case of remote access, thesis also provides complete high level design, which solves all discovered weaknesses and upgrades the security of ATLAS environment. In case of data management, this thesis was not fully successful. The analysis successfully discovered possible improvements in the system but it was not able to offer complete suggestion of remediation of those weak spots. This is caused by absence of required tool or technique which could be implemented into the environment.

There are several ways in which this thesis could be further broadened. The obvious one is to start working on detailed design of suggested solutions and start implementing them. Another less direct possibility is to continue with testing of the ATLAS environment and start suggesting and designing improvements. One of the possibilities, which would have great impact, is to analyze ATLAS infrastructure and accounts, in order to further extend the account separation and design tiering in the environment.

Appendix A

Acronyms

AAL authentication assurance level.

ABAC Attribute-based access control.

AD Active Directory.

CA certification authority.

CERN European Laboratory for Particle Physics.

CRL certificate revocation lists.

DNS domain name service.

DOS denial of service.

DRM data rights management.

EDG European Data Grid.

EGEE Enabling Grids for E-science.

EU European Union.

GUI graphical user interface.

HSM Hardware security module.

IBAC Identity-based access control.

KDC key distribution center.

LCG LHC Computing Grid.

LHC Large Hadron Collider.

MC Monte Carlo.

MFA multi factor authentication.

OSG Open Science Grid.

PAM privilege access management.

PAW privileged access workstation.

RAdAC risk-adaptable access control.

RBAC Role-based access control.

SIEM security information and event management.

SSH secure shell.

SSO single sing on.

TGS ticket-granting service.

TGT ticket-getting ticket.

TTL time to live.

VO Virtual Organization.

VPN virtual private network.

WLCG Worldwide LHC Computing Grid.

Bibliography

1. SHIREY, R. Internet Security Glossary, Version 2. In: [online]. 2007 [visited on 2021-02-13]. Available from: `https://tools.ietf.org/html/rfc4949`.
2. LÓRENCZ, R. Základní pojmy v kryptologii, substituční, blokové a transpoziční šifry. In: [online]. 2020 [visited on 2021-02-13]. Available from: `https://courses.fit.cvut.cz/BI-BEZ/media/bez1.pdf`.
3. LÓRENCZ, R. Exponenciální šifra, zřízení společného klíče a problém diskrétného logaritmu. In: [online]. 2020 [visited on 2021-02-13]. Available from: `https://courses.fit.cvut.cz/BI-BEZ/media/bez2.pdf`.
4. PETERKA, J. *Báječný svět elektronického podpisu* [book]. CZ. NIC, 2011. ISBN 978-80-904248-3-8.
5. COOPER, D.; SANTESSON, S.; FARRELL, S.; BOEYEN, S.; HOUSLEY, R.; POLK, W. Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile. In: [online]. 2008 [visited on 2021-02-13]. Available from: `https://tools.ietf.org/html/rfc5280`.
6. GRASSI, P. A. et al. *Digital Identity Guidelines: Authentication and Lifecycle Management* [online]. National Institute of Standards and Technology, 2020 [visited on 2021-02-13]. Available from DOI: `https://doi.org/10.6028/NIST.SP.800-63b`.
7. RICCIARDI, F. *Kerberos Protocol Tutorial* [online]. 2007 [visited on 2021-02-13]. Available from: `https://www.kerberos.org/software/tutorial.html`.
8. PRIVILEGE MANAGEMENT CONFERENCE COLLABORATION TEAM. *NISTIR 7657: A Report on the Privilege (Access) Management Workshop* [online]. National Institute of Standards and Technology, 2010 [visited on 2021-02-14]. Available from DOI: `https://doi.org/10.6028/NIST.IR.7665`.

9. COYLE, K. *The Technology of Rights: Digital Rights Management* [online]. 2003 [visited on 2021-02-14]. Available from: https://kcoyle.net/drm_basics.pdf.

10. SOUPPAYA, M.; SCARFONE, K. *NIST Special Publication 800-46: Guide to Enterprise Telework, Remote Access, and Bring Your Own Device (BYOD) Security* [online]. National Institute of Standards and Technology, 2016 [visited on 2021-02-14]. Available from DOI: https://doi.org/10.6028/NIST.SP.800-46r2.

11. CENTER, Australian Cyber Security. *Secure Administration* [online]. 2020 [visited on 2021-02-14]. Available from: https://www.cyber.gov.au/sites/default/files/2020-06/PROTECT%20-%20Secure%20Administration%20%28June%202020%29.pdf.

12. *CNSSI No. 4009: Committee on National Security Systems (CNSS) Glossary* [online]. 2015 [visited on 2021-02-14]. Available from: https://rmf.org/wp-content/uploads/2017/10/CNSSI-4009.pdf.

13. CHARPENTIER, P. *LHC Computing: past, present and future* [online]. 2019 [visited on 2021-02-14]. Available from DOI: https://doi.org/10.1051/epjconf/201921409009.

14. *User Documentation* [online]. 2017 [visited on 2021-02-14]. Available from: https://wiki.egi.eu/wiki/User_Documentation.

15. POLINI, A.; CERRI, A. *General Introduction for all shifters* [online]. 2014 [visited on 2021-02-14]. Available from: https://atlasop.cern.ch/twiki/pub/Main/M8ShiftTraining/Shift-General.pdf.

16. POLINI, A.; CERRI, A. *Shift Leader Training Part 1* [online]. 2015 [visited on 2021-02-14]. Available from: https://atlasop.cern.ch/twiki/pub/Main/M8ShiftTraining/Shift-Leader-Part1.pdf.

17. POLINI, A.; CERRI, A. *Shift Leader Training Part 2* [online]. 2015 [visited on 2021-02-14]. Available from: https://atlasop.cern.ch/twiki/pub/Main/M8ShiftTraining/Shift-Leader-Part2.pdf.

18. BERLINGEN, J. M. *Triggering in the ATLAS Experiment* [online]. Geneva, 2020 [visited on 2021-02-14]. Available from: https://cds.cern.ch/record/2723014. Technical report. CERN. ICHEP 2020 slides, started 28th July.

19. COLLABORATION, ATLAS. *ATLAS Publication Policy* [online]. Geneva, 2006 [visited on 2021-02-14]. Available from: https://cds.cern.ch/record/2702494. Technical report. CERN.

20. LLOYD, S. *Starting on the Grid* [online]. 2018 [visited on 2021-02-14]. Available from: https://twiki.cern.ch/twiki/bin/view/AtlasComputing/WorkBookStartingGrid.

21. LLOYD, S. *Getting an Account* [online]. 2018 [visited on 2021-02-14]. Available from: `https://twiki.cern.ch/twiki/bin/view/AtlasComputing/WorkBookGetAccount`.

22. WEISL, J.; SOPCZAK, A. *Personal consultation about ATLAS environment*. Prague, 2020.