Trust Management in Space Information Networks

Anders Fongen
Norwegian Defence University College, Cyber Defence Academy (FHS/CISK)
Lillehammer, Norway
Email: anders@fongen.no

***Abstract*—The concept of a Space Information Network (SIN) is evolving from a satellite transport infrastructure towards a provider of a range of services, including even Application-as-a-Service (AaaS). Client endpoints connected to a SIN will invoke services in other connected endpoints, as well as services inside the SIN itself. Interactions taking place between clients and SIN components will create trust relations that must be protected from the usual threats. Traditional cryptographic protocols can offer adequate protection from some threats, but the particular conditions of a satellite network requires modifications of the methods used for authorization control and key management. The amount of connectivity and transport capacity required by a traditional Public Key Infrastructure (PKI) configuration causes excessive use of SIN resources, and a modified approach to key deployment, credential validation and authorization control should be investigated.**

Keywords*—*LEO satellites; trust management; space information networks; AaaS in space

I. Introduction

The term *satellite networks* indicates the evolution of satellites from being radio mirrors to form complex infrastructures where the spacecrafts cooperate for the provisioning of communication services. Satellite networks for communication services have been in operation for three decades and have proven the feasibility of their operation, capacity and utility. We foresee the further evolution of satellite networks into the *Application-as-a-Service* (AaaS) domain, where the network not only provides communication services, but also different kinds of discovery services, collaborative services and even platforms for general AaaS. The descriptive term for this evolving concept is *Space Information Networks* (SIN). Not only will a SIN provide global coverage, but also a very low Round Trip Time (RTT). A satellite at 300 km altitude can offer an RTT as low as 2 ms, much less than any terrestrial network path.

The evolution presented in the above paragraph creates service endpoints inside the network elements of the SIN, representing high value for both providers and customers, so trust management must be in place not only between client endpoints, but also inside the SIN infrastructure, as services in satellites are invoked from other satellites and client endpoints. Existing technology for authentication and authorization control may not be well suited for the particular properties of a SIN infrastructure, which this paper aims to address.

The illustration in Figure 1 shows the endpoints involved in transactions in or trough a SIN: The *Client Endpoints* (CE) are computers connected to the SIN (blue lines). A CE can both have client and server roles, but they are still clients to the SIN services. The service endpoints in satellites are called *Satellite Endpoints* (SE) and may be invoked from CEs as well as other SEs. There are a number of terrestrial endpoints called *Ground Stations* (GS), used by the satellites for communication with the Internet. Services offered by GS are never invoked from CE, only from SE. Red lines in the figure indicate intra-SIN communication endpoints not addressable for CE use.

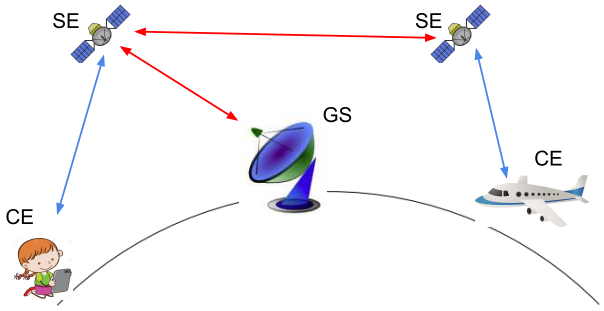

Figure 1. Service endpoint and links which forms the structure of a SIN

.

The general architecture principle of an AaaS oriented SIN has been published in a previous article [1], where a number of future research problems were presented. In the present paper, a model for SIN trust management will be described in some detail. The general principles of the proposed trust management architecture have originally been developed with tactical military networks in mind [2], and have been modified to match the properties of a satellite network.

A key property of Low Earth Orbit (LEO) satellites is the long idle periods as they fly over inhabited areas, and the predictability of the bursts of requests they receive as they fly over densely populated areas. Non-interactive tasks can thus be scheduled to idle periods, where data stores can be replicated, software updated, etc. Intelligent replication of frequently used resources can contribute to reduced latency and efficient use of infrastructure capacity. [1].

The contribution of this paper is a model for key management, authentication and authorization control using protocols well suited for the particular properties of a SIN. The identification of *Delay Tolerant* operations in credential management that can be scheduled to idle periods is essential in this respect.

The remainder of the paper is organized as follows: In Section II, a short survey of relevant research is presented. Section III identifies the shortcomings of the PKI design.

 ISBN: 978-1-61208-919-5

Section IV presents the author's alternative to X.509, the *Identity Statement*, and how its properties better serve the purpose of trust management and protected service invocation in a SIN. Section V summarizes the arguments of this paper and identifies future research activities.

II. Related research

The term *Space Information Network* (SIN) has been used to describe networks of satellites and high altitude aircrafts (drones, balloons) with different service levels. Existing satellite networks like Iridium and the upcoming Starlink [3] offer only communication services, the latter on a very large scale and with high bandwidth. A number of authors have proposed "Cloud Computing in Space" through the addition of larger satellites with sufficient energy and computing resources for taking on these tasks [4] [5].

In order to improve the communication capacity of SIN units, lots of research has gone into the development of antennas for spatial multiplexing (Space-Division Multiple Access, SDMA), beamforming, non-orthogonal multiple access, optical communication links, etc. [6] [7]

The proposals made in this position paper will not deal with technical details in the communication technology, but rather view the SIN as a distributed system which borrows its analysis and solutions from the field of distributed computing. The author is not aware of other efforts to investigate trust management and protection mechanisms specifically for a SIN. Efforts on trust management are made in related areas, as in Mobile and Distributed Systems [2], and in the area of Internet of Things (IoT). IoT systems seem to show little interest for traditional PKI, but rather look to the use of Blockchains. In [8], Blockchains are proposed as the distribution method for tamper-proof trust variables, which are formed through consensus processes and transitive trust. Given that Blockchains have scalability problems, [9] proposes a variant called Holochain, with better scalability properties since the distribution patterns are limited.

Proposals based on Blockchain/Holochain for trust management seem to overlook the importance of the trust chain which binds the technological domain to the managerial domain through cryptographic protocols, and the complexity of the resulting *key management*. Which is why these efforts are not used as a basis for this paper.

III. Public Key Crypto and Infrastructure

The reader is assumed to be familiar with the fundamental principles of public key crypto, digital signatures, cryptographic hash functions and Public Key Infrastructure (PKI).

The PKI services can be divided in two categories:

1) Creation and deployment of key pairs and certificates
2) Assistance in the certificate validation process.

Operation (1) takes place for each End Entity (EE) after the existing certificate expires, while operation (2) takes place at short intervals or even every time a certificate is validated. It is the task of certificate validation which demands the highest connectivity and network capacity, which is why it is of interest for operation in a SIN.

A. Certificate revocation

The decision that a certificate should no longer be validated is called *revocation*, and is made by the Certificate Authority (CA) and announced to the community in a variety of ways. A common method is to offer an interactive service through which EE can check the revocation status of a certificate by using the *Online Certificate Status Protocol* (OCSP) protocol. Another method is to disseminate a *revocation list* of certificates that are revoked but not yet expired. Experience indicates that approx. 10% of the certificate population is revoked and represented on a revocation list [10] through entries of (typical number) 37 bytes each.

The use of revocation lists has never been a good idea, and although attempts have been made to distribute delta lists and fragmented lists, the required network capacity for their dissemination is massive [11]. Besides, revocation lists raise lots of dilemmas in situations where the dissemination fails, which is considered to be out of scope for this paper [12].

B. Authorization control through certificates

Certificates facilitate the authentication phase through binding a transaction or an object to an identifier. It does not indicate the *authorization* of the corresponding entity. Authorization control involves a new set of data sources and protocols for their distribution. Although standards have been published for its interoperability, e.g., XACML [13], they are not widely used. Most vendors offer their own proprietary solution.

In order to avoid the extra cost associated with separate authorization control, many systems choose to confuse authorization with authentication, and assume any valid certificate to be a token for authorization. This is a mistake, which greatly increases the need for revocation, since any changes in the authorizations of an entity requires a certificate to be revoked and a new certificate issued.

In a *constrained network*[14], both authentication and authorization control should be done using one set of data objects and protocols. The most popular standard format for certificates, the X.509, does not lend itself well to this combination, for which reason a different data structure is proposed: The *Identity Statement* (IdS).

IV. The Identity Statement

For the purpose of authentication and authorization control in a constrained network, the protocols in use should have as few round-trips as possible with the smallest messages possible. For this purpose, the object class *Identity Statement* (IdS) has been constructed. It has many similarities with an X.509 certificate, but is simpler, and a block of named variables (name-value pairs) has been added to support Attribute Based Authorization Control (ABAC) operations. Its elements are:

- Identifier of subject, RFC-822 format address
- Public key

 ISBN: 978-1-61208-919-5

- Validity period
- Authorization attributes
- Issuer's Distinguished Name (X.500 form)
- Issuer's signature
- Room for cross-CoI extensions (described later)

The public key in the IdS can be used both for signature verification (during authentication) and encryption, but not for issuing new Identity Statements. There is no *keyUsage*-element, which means that keys can serve any purpose. As in a PKI, the trust chain depends on a small number of Trust Anchors, called *Identity Providers* (IdP). Their X.500 DN and digital signature are stored in each IdS and used for IdS validation. The group of clients which have the same IdP as their trust anchor is called a *Community of Interest* (CoI).

There is no revocation operation in this architecture. The IdS is irrevocably valid until it expires, before which it is re-issued unless it is invalidated in the mean time. The validity time may be set so short that it matches the *revocation latency* associated with revocation list (typically a small number of hours). The dissemination of re-issued IdS takes opp much less capacity than a similar arrangement based on revocation lists.

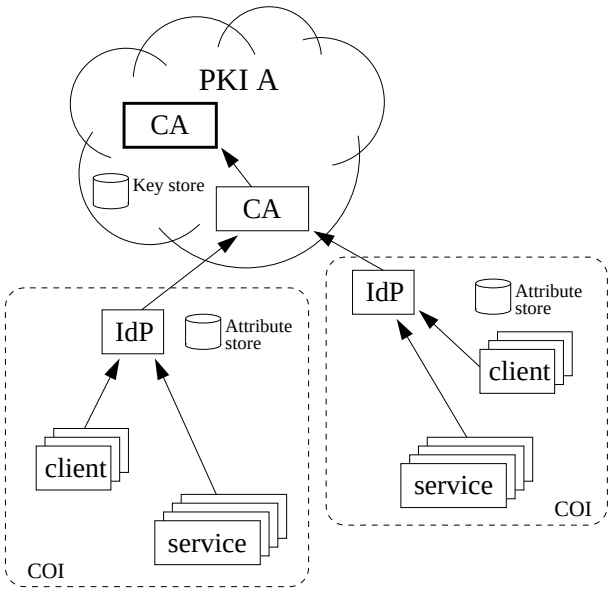

Figure 2. The functional components of trust management. The IdP serves one single CoI. Keys are issued by a PKI, attributes by the IdP.

A. Issuing Identity Statements

The authority which issues Identity Statements is the Identity Provider (IdP). The structure of the issuing service is shown in Figure 2. The IdP keeps all EE information in a database (possibly gets it from a traditional PKI) and provides signed IdS at anyone' request through a simple HTTP interface. The IdS is a public object so no caller privileges is needed. The public key of the IdP must be installed and trusted by every EE in order for them to validate an IdS.

If the IdP receives an IdS issued by a different IdP, the IdP will issue a *Guest Identity Statement* with the same content and a selection of its authorization attributes, based on a *trust relation* between the two IdPs. This is a way for guest clients from a different CoI to invoke services in this domain. This approach to cross-CoI validation is vastly more efficient and secure than the cross-certificate approach proposed by the traditional PKI.

B. Dissemination of re-issued Identity Statements

An endpoint (CE or SE) must possess a valid IdS of the corresponding party in order to validate an authentication request, cf. Section IV-C. Normally, it would be the responsibility of the requesting part to enclose a valid IdS with the request message, but several other communication patterns are possible. The validating party may store the IdS from earlier transactions, or may request it directly from the IdP service point.

Please keep in mind that all authentication operations should be *mutual*, i.e., both parties authenticate to the other, and both must have a valid IdS representing the other party at the time of authentication.

Since IdS are never revoked, sound practice for the IdP is to give them a short expiration time and renew them on demand. Anyone possessing an IdS will know the time for its expiration and can plan a suitable moment for its renewal. For this reason, the dissemination may be regarded as a *delay tolerant operation*, which takes place in a relaxed manner when the satellites are in a favorable position for the operation. The satellite can receive the IdS when it is directly communicating with a Ground Station (GS), and pass it on to the CE later when it is within range. In this way, the delay tolerant properties of the operation may allow for the satellite to be used as a *courier* rather than consuming infrastructure capacity.

For an IdS which represents the service endpoint in a satellite, the problem is simple. As the expiration time for the existing IdS is due, the satellite requests a new from the next GS in range.

For CEs, the courier approach raises interesting questions: (1) which satellite(s) should be chosen for the courier task, and (2) when is the CE in operation and ready to receive the IdS? The following observations apply for the analysis of possible solutions:

1) The CE has only one connection point, which is a satellite. The IdS may as well be stored in the satellite as in the terrestrial endpoint. Besides, the satellite has a shorter path to the IdP and higher communication capacity. The IdS will be a part of the client state during handover to trailing satellites before being discarded. A complicating factor for this arrangement to work is that the SE need to engage in the authentication protocol and inject the IdS into the message stream when needed.
2) If the CE is authenticating with an Internet endpoint, the other endpoint has the most network capacity to its disposal. It may as well acquire the IdS for the CE by itself, and cache it for subsequent invocations.

 ISBN: 978-1-61208-919-5

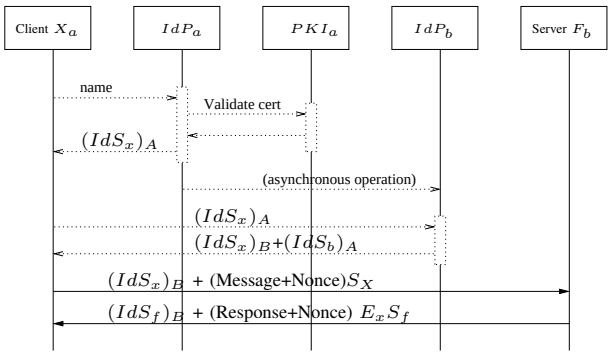

Figure 3. Trust management protocols for IdS issue and service invocation in a cross-CoI environment.

3) The IdS could be replicated on a subset of satellites, so that the CE may connect to one of them within a given time period (e.g., 60 minutes) to find a renewed IdS. With a handover frequency of 10 minutes (typical number) at least every 6th satellite passing over the CE should be able to offer the IdS. This fraction can be lower if the location of the CE is known or guessed, and the validity period of the IdS is less than a full orbital period. One can also take advantage of the fact that a southbound satellite pass will be northbound 12 hours later. There is a trade-off between the number of satellites involved and the operating demands on the CE, e.g., if the CE always has to be connected to the SIN.

As a fallback option, the endpoint may invoke the communication service to obtain an IdS from the IdP service point. Under the proposed scheme for IdS dissemination, this service is likely to be the choice when the CE computer is started and used immediately, if it cannot wait for the next pass of a courier satellite.

C. Invoking services with IdS

The protocol for invoking a service should provide mutual trust establishment through a minimum number of messages. In the simplest scenario, the requester/client will send its IdS together with the request message and a nonce, signed by its private key. The responder/server will validate the IdS, verify the signature and execute the service. The response message will include the server's IdS and the service response and the nonce, encrypted with the client's public key and signed with server's private key.

Figure 3 illustrates a *cross-CoI* service invocation, which involves IdS issued by two IdPs, a guest IdS for client X issued by IdP_b, a cross-CoI $(IdS_b)_a$ for IdP_b issued by IdP_a for validation of the server's IdS by the client. Apart from these extra data elements, the cross-CoI invocation remains essentially similar to the base case, and there is no need for revocation status from foreign CoIs, which would otherwise complicate the validation of the guest IdS. The initial invocation of the IdP services and the enclosure of IdS in the service invocation messages are not strictly necessary since they may be cached in the parties from preceding operations.

V. Conclusion

This paper describes the trust management components of an ongoing effort to outline the design of a Space Information Network with application service capabilities (AaaS). Its main focus is to preserve low latency through prudent protocols and data structures, as well as room for any number of credential-issuing authorities (called *Identity Providers*, IdP).

Why is the proposed trust management essential for the SIN operation? Because it allows cross-CoI service invocations to take place in a minimum of round trips and with minimal message size, allowing the SIN to offer services with unprecedented low latency, which is the most important motivating property for its design.

Other revocation free schemes could possibly work, like replacing the short-lived IdS with a combination of X.509 certificates and an OCSP response message which attests the validity of the certificate for a short period of time. This approach does not, however, lend itself well to the inclusion of authorization information in the trust protocols. Besides, the validation of an X.509 certificate involves a large number of poorly understood variables, which is often seen to create errors, ambiguities and interoperability problems.

Issuing and dissemination of IdS remains an unsolved problem though, which should take place in a *delay tolerant* manner to exploit the frequent idle period of satellites as they fly over inhabited areas. A simulation model is under construction for the study of possible solutions.

References

[1] A. Fongen, "Application services in space information networks," in *CYBER 2021*. Barcelona, Spain: IARIA, Oct 2021, pp. 113–117.

[2] ——, "Federated identity management in a tactical multi-domain network," *Int. Journal on Advances in Systems and Measurements*, vol. Vol.4, no 3&4, pp. 157–167, 2011.

[3] "Starlink web site," https://www.starlink.com/, [Online; accessed 19-Oct-2021].

[4] S. Briatore, N. Garzaniti, and A. Golkar, "Towards the internet for space: Bringing cloud computing to space systems," in *36th International Communications Satellite Systems Conference (ICSSC 2018)*, 2018, pp. 1–5.

[5] S. Cao *et al.*, "Space-based cloud-fog computing architecture and its applications," in *2019 IEEE World Congress on Services (SERVICES)*, vol. 2642-939X, 2019, pp. 166–171.

[6] X. Zhang, L. Zhu, T. Li, Y. Xia, and W. Zhuang, "Multiple-user transmission in space information networks: Architecture and key techniques," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 17–23, 2019.

[7] Y. Su *et al.*, "Broadband leo satellite communications: Architectures and key technologies," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 55–61, 2019.

[8] A. Lahbib, K. Toumi, A. Laouiti, A. Laube, and S. Martin, "Blockchain based trust management mechanism for iot," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, 2019, pp. 1–8.

[9] R. T. Frahat, M. M. Monowar, and S. M. Buhari, "Secure and scalable trust management model for iot p2p network," in *2019 2nd International Conference on Computer Applications Information Security (ICCAIS)*, 2019, pp. 1–6.

[10] S. Berkovits, S. Chokhani, J. Furlong, J. Geiter, and J. Guild, "Public key infrastructure study: Final report," *Produced by MITRE Corporation for NIST*, April 1995.

[11] A. Fongen, "Optimization of a public key infrastructure," in *IEEE MILCOM*, Baltimore, MD, USA, Nov 2011, pp. 1440–1447.

 ISBN: 978-1-61208-919-5

[12] R. L. Rivest, “Can we eliminate certificate revocation lists?” in *Financial Cryptography*, R. Hirchfeld, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 178–183.

[13] “OASIS eXtensible Access Control Markup Language,” https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml, [Online; accessed 19-Oct-2021].

[14] C. Bormann, M. Ersue, and A. Keränen, “Terminology for Constrained-Node Networks,” RFC 7228, May 2014. [Online]. Available: https://rfc-editor.org/rfc/rfc7228.txt