International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:14, No:11, 2020

Decentralised Edge Authentication in the Industrial Enterprise IoT Space

C. P. Autry, A.W. Roscoe

***Abstract*—**Authentication protocols based on public key infrastructure (PKI) and trusted third party (TTP) are no longer adequate for industrial scale IoT networks thanks to issues such as low compute and power availability, the use of widely distributed and commercial off-the-shelf (COTS) systems, and the increasingly sophisticated attackers and attacks we now have to counter. For example, there is increasing concern about nation-state-based interference and future quantum computing capability. We have examined this space from first principles and have developed several approaches to group and point-to-point authentication for IoT that do not depend on the use of a centralised client-server model. We emphasise the use of quantum resistant primitives such as strong cryptographic hashing and the use multi-factor authentication.

***Keywords*—**Authentication, enterprise IoT cybersecurity, public key infrastructure, trusted third party.

I. Introduction

THE PKI model of authentication was designed with a simple world in mind. Public key cryptography, the basis of PKIs, was developed in GCHQ and independently in the non-classified world during the 1970's, and the idea of a PKI developed quickly after that [1].

PKI is ideally suited to a world where there are relatively few computers, and where if *A* wants to talk to *B* she will know *B*'s name in a way that clearly distinguishes *B* from all other players. There is a TTP called a certification authority (CA) who has carefully checked *B*'s identity and issued him with a key certificate for his public key pk_B. When *A* sees this certificate, she knows either how to send him a message only he can decrypt or can verify a signature that proves that *B* was really the author of something. The security chain is complete because *A* already knows the CA's public key, which allows her to check that the certificate is valid. However, in the Internet of Things (IoT) there are billions of nodes which, at least from the point of view of the world at large, lack any commonly understood sense of identity. And even if they did, the cost of creating and maintaining a PKI is prohibitive. In all likelihood a sensor only gains a meaningful identity when it has a known owner, hub and position. However, establishing a secure connection with a CA to get a certificate when it is in this position is precarious indeed. In other words, we believe that this model is not suited to IoT. IoT is one of a number of spheres where understanding the *context* of a node (its position, neighbours etc.) is as least as important as its name [2].

In practice CAs consist of a number of layers meaning that anyone checking certificates needs to check multiple signatures, including some strong and therefore slow ones. The more structured and larger a PKI is, the more layers of CAs are typically required. The costs of running a PKI model, particularly when we have multi-layered CAs, is prohibitive in computing power and battery consumption because of the complex calculations required. PKIs generally have other functions such as giving time-bounded certificates. To exploit this, we would need to work out how to apply it with already deployed sensors. We need a new security model which maps naturally onto IoT.

To date, it is highly unlikely any viable alternative protocol has been proposed beyond PKI/TTP to function in large scale multiparty enterprise IoT environments beyond the gateway level of authentication. Relying on gateway authentication to secure sensor and sensor hubs, in other words, the edge of complex industrial IoT systems, is fraught with vulnerabilities including but not limited to, brute force attack, man-in-the middle, and severe post-quantum exposure. If a gateway is exposed, so too are all the components that rely on the gateway for connectivity, routing and often functionality.

There are currently improvements to the original PKI protocols [3] however, all rely on centralised client-server configurations, static databases of public and private keys and require significant CPU capacity, and none have been designed with post-quantum world in mind. This is significant as it is estimated the current quantum computing capability allows for the decryption of any current public/private key exchange in milliseconds [4]

Authentication is fundamentally about trust between two parties that wish to share information between them. The parties' required belief that each is what it says it is, and both parties believe this, is fundamental to the process. Authentication precedes authorisation and often the two go hand-in-hand. Using a number of essential components, we believe that it is possible for the edge of complex IoT systems to in effect bootstrap authentication between assets in real-time without the need for private/public key exchange.

II. Methodology

An approach to quantum resistant authentication of assets at the edges of large scale IoT system, not dependent on a static database of third party certificates held by a third party using the following components, is proposed.

C.P. Autry is CEO of Iothic, Ltd, London, United Kingdom (e-mail: cpautry@iothic.io).

A.W. Roscoe, FREng is a professor and former head of department at Oxford University Department of Computer Science, Parks Road, Oxford OX1 3QD and Chief Science Officer (CSO) at Iothic (e-mail awroscoe@iothic.io).

A. Quantum Resistant Primitives

For authentication alone, it is possible to restrict cryptography to symmetric ciphers such as the advanced encryption standard (AES), cryptographic hashing such as secure hash algorithm 3 (SHA3) and perhaps short-life quantum resistant (e.g. lattice-based) signature algorithms. We therefore can avoid both the use of cryptography vulnerable to Shor's algorithm such as Diffie-Hellman and ECC, and long-term dependence on cryptographic algorithms whose security may not last or which are expensive to run or store. We have very efficient proposals for this.

B. Cryptographic Hashes

A *cryptographic hash function* maps arbitrary length inputs to fixed length outputs. They create high quality digests of their inputs from which (i) it is next to impossible to tell what the input, or indeed any other input producing the same output, might be and (ii) it is in any practical sense impossible to find a pair of different values mapping to the same output (a *collision*).

For long term authentication we use an efficient method based on cryptographic hashing. Unlike novel post quantum algorithms, this is well understood type of function. Subject to doubling its output length, it is widely believed to be quantum resistant, as are symmetric encryption algorithms such as AES. The large internal state of SHA3 makes it an excellent candidate for complex multi-node environments.

It is possible to base authentication, as opposed to encryption, solely on cryptographic hashing, and indeed hash-based signature schemes such as Lamport's [5] have been known as long as ones based on asymmetric encryption. We choose to base initial authentication of nodes on hashes of shared data and data stream authentication on a much more efficient hash-based scheme than Lamport's or any other we are aware of.

C. Independent Executables at the Node Level

It is wholly possible given the minimal computational power at the edge of IoT systems, namely sensors and sensor hubs, to instruct these assets to perform certain low-level commands. Moreover, many of these sensors themselves, but most certainly sensor hubs, operate on multiple wireless channels [6].

D. Commonality between Nodes and Proxies

In order for parties to start an authentication process between them, some form of trusted commonality must exist. This requirement has traditionally been filled by TTP. However, this model requires a static database of certificates and a centralised means of communication of access in order for both parties to communicate with the TTP, and as previously mentioned is prohibitive in terms of both cost and CPU power required. And as described above, it ignores the fact that the identities that are being verified may be hard to establish and propagate: in practice these are unlikely to be known much beyond the hub that controls the sensor.

Our model involves using local knowledge of nodes gained through provisioning and interacting with proxies. This is both appropriate for bootstrapping security between a sensor and the hub and its peers in its immediate network, and we can then use networks of established keys and trust to extending this security, enabling remote nodes to connect securely. We have an adaptable model of authentication which exploits both these and other factors to bootstrap and re-confirm security.

III. Proof of Concept

To ascertain the possibility of bootstrapping authentication between nodes or groups of nodes in a point-to-point decentralized environment, we have built and designed a proof of concept using Unix, its core language C, and C++ in combination to test the validity of our claims. Both C and C++ are commonly used lower level programming languages dedicated to instruct real world assets such as routers, gateways and other various digital industrial assets. To test this modality, we provisioned like assets (*A*,*B*) with common groupings that match another, third asset, *C*. In other words $A \equiv B \equiv C$. In our initial tests, we want *A* and *B* to authenticate with one another and to use a third asset, *C*, to generate a hash that is unique to the initial *A*, *B* coupling. In order for this to occur, all assets in the group must know that it is part of such said group, and can identify and communicate authentically with others in the group. To achieve this, we can assign any number of identical values bitstrings or floating point numbers to all assets of the same group that are recognised by other member of that group and not initially known to outsiders. These might be the coordinates of the assets or their hub, statistics of interactions, purely arbitrary values, or whole or partial combinations of them all. Note that as well as being initialised with the values appropriate to themselves and other members of their group, nodes may well be able to *sense* some of them.

Additionally, we have assigned assets *A* and *B* with any number of floating-point values, and a number of alphanumeric descriptors that are unique to each asset. From these values, we are able generate a 512-bit primitive that is unique to each asset, A and B as in Table I.

TABLE I
SHA 3 Conversions of Real and Virtual Asset Properties

Asset	Descriptor	SHA3 Conversion
A	Red,23887A,1.443, AdF002132, Model213086	b5d84b8cf28a5e7182a55aca4954f26625 94225fbdee0da85fb817262d4d5002de9 c8b9ad5a67adb543a18aa64d000ed938d 67f3009a0ae2b4e001b77855d808
B	White4,-2.338af, January18:20122,Ser1011342284aEd,PortFor	ffb0eb3d85badcebb792bf9f99b5e0d5c7 442e68462f8d0de56644220097449072f 7886fba99f515daa16169354d6fa5d75cc e4da7f67cfa95d6f84152513e01

Our main objective was to use the unique properties of each asset, *A* and *B*, to generate a third unique digest created at *C*, and then have this unique value passed back to both *A* and *B* in order to complete the authentication process. For example, we can hash the concatenation of the above two values combined with some structural information to generate a third digest *ee52875afc651604cd757f1f27fabc34883fd34c3b6bce0b5d1dd*

6a8b544d332e488f09b8a50dbfd471b74fca6c644aa1318430a6a3206cc95277b4b9a0fc2b0. Similarly, we can use any or all of an asset's SHA3 conversion, in combination with any other part of any or all of the second asset's SHA3 conversion, to generate a third digest. This can change with each executable, creating substantial entropy and rendering the final digest impossible to predict beforehand. The final digest can then be passed back to both, A and B simultaneously, the expected hashsum confirmed by both parties independently, and the authentication is allowed to complete.

We have designed the test model to be recorded and tracked into the blockchain using both the initial and final SHA 3 digests. This allows for rigourous and precise reporting of all attempted, failed and successfully authentication attempts between nodes and groups of nodes.

IV. Proposed Model

Our protocol seeks to use a wide variety of information to authenticate networks to make it essentially impossible to hack successfully. The aim is to make it impossible to successfully hack just one part of a network because the checks and rechecks would find inconsistencies. Due to the static nature of the PKI model, brute force attacks at the gateway level of IoT architecture are currently commonplace thereby exposing the functional operations levels below the gateway, in other words at the sensor, sensor hub level in IoT systems. Any certificate based authentication becomes even less reliable as future quantum processing comes online. By authenticating at the edge of complex IoT systems, it becomes nearly impossible to hijack enough of the system to control or impede overall functionality.

We propose a model which can either be used to bootstrap an entire network or augment an existing one. Further, unlike standard authentication we propose a one-time, one-off, as needed, node to node bootstrap authentication model which occurs as required and is different each time.

The core of our model is a protocol in which provisioning is used to establish a base level of authentication between nodes. Each node is authenticated to others that share knowledge of its initial provisioning. Such knowledge may include AES keys, in which case the exchanges needed for this authentication can be encrypted under it. However, to be safe we either have to be sure that each key is individual to a node and shared only by its hub, in which case the node can only be authenticated to its hub, or we need to be sure that no one else who has the key has been corrupted. Thus such keys can be useful but are only part of the story.

It is entirely possible to use the same provisioning information in multiple authentications with different parties without giving it away. Instead of two nodes each confirming that they share information directly, each time a comparison needs to be done, a hash or partial hash of it combined with a freshly chosen and shared nonce.

We note also that if A is a node then different nodes B,C,… may know different subsets of its provisioning information and therefore use different comparisons to authenticate it, perhaps separately judging the entropy strength of each.

Based on the provisioning that is shared, a node may judge that it should have some physical evidence of a second one's presence. Frequently this will manifest itself as a channel (e.g. a short range radio one). Detection of appropriate information provides a necessary part of provisioning based authentication where it should be present. For example if node B does not see a channel from node A when it knows it should be there, then authentication will fail.

Where such alternate channels are present then, using them for authentication gives more confidence than getting the same information over a channel that might have come from anywhere. In other words, it is wholly possible to use a multichannel approach, one for recognition between nodes, and the other to pass proprietary hash exchanges between them.

Once either the complete network or part of it has been initially authenticated then we reconfirm the agreement on the overall consistency of agreement on the provisioning over secondary channels provided by either alternative (provisionally authenticated or higher physical confidence) channels. Furthermore all authentications are confirmed round the entire network. We have efficient and strong protocols to achieve this. Thus, for example, any failure to observe a local channel that is expected between B and C will propagate to all other A. Should the expected value not occur, no authentication will occur. In order to be completely authenticated, all nodes will need to know that all others are happy. They are, because of the way we have constructed this, using all the evidence at their disposal. We therefore believe that it is much more secure than an authentication protocol that only examines one dimension of a system controlled by a third party.

An important topic for future research is rating the security level (something like a confidence level) of a particular authentication. After all, we will get greater confidence from a case whether there are many interconnections and pieces of checkable physical evidence than one where there are few. In the case of IoT, if the many components at the edge were able to authenticate independently of one another, the likelihood of an entire system becoming compromised, as is currently the case, is much less probable. This thus gives us, much as with PKI, a general model of authentication, but one which we believe is much better suited to complex ever changing IoT system architectures. We intend to implement it in a general sense and then produce specialised versions for different domains. As not all IoT edge components are created equal, contingencies must be made available for any number of instances. Moreover, we must allow for the sudden addition or transformation of physical assets without compromising an interconnected systems' integrity.

At present, we are rigorously testing our protocol in a range of interconnected physical assets, with a wide range of CPU capability, and under a large range of conditions to ascertain reliability and integrity. We are conducting these trials in both multiple channel wireless and combination wireless and wired environments.

V. Conclusion

The main objective of our trial was to create a point-to-point decentralized authentication process which is new each time two assets or groups of assets attempt to authenticate with one another. Moreover, IoT assets generally need only communicate to one another on by-need basis and have no reason to maintain a permanent connection after the initial exchange of information. After all, we should be able to bring in a new asset and augment the existing network by it without compromise. This is advantageous in that if there is nothing to authenticate, disruption of authentication is impossible. One of the main weaknesses of the current authentication model is indeed the static, always connected nature. Further, if simple computation and instruction sets can be sent out to the edge of complex IoT systems, it is wholly possible to devise a protocol better suited for these environments that do not rely on a centralized, static database of public keys required to complete authentication between assets. It is wholly possible to generate on-the-fly private key equivalents between nodes using little or no CPU capability.

The introduction of proximity and location time based provisioning into a model of authentication may also prove advantageous. A basic model of IoT environment requires an imminent local reason to communicate with some party other than itself at a particular place and time. While it may be simple to establish a trust model between nodes in an environment which are wholly controlled by one entity, this is not true of a dynamic IoT structure where many parties communicate with the assets of many other parties. Indeed, if the concept of authentication was approached based only on functionality required at a specific time, based on certain criteria being met, location, for example, and is different each time it occurs, this leaves little to be hacked. If two parties connect momentarily or not at all, this is no chance for man-in-the middle or other insertions.

Even at the level of cryptography (post quantum or otherwise) our protocol is enormously more economical than PKI-based ones. We also believe it is more appropriate because it builds on the existing structures of IoT rather than bolting on structures that were devised for a permanently connected client-server environment. The decentralised, unpredictable nature of a truly dynamic IoT environment can be highly advantageous and leveraged accordingly. And the associated security protocols can readily reflect this. Our current model of point-to-point, dynamically generated, decentralised, bootstrapped authentication is designed specifically for the edge of complex IoT systems.

References

[1] Ellis J.H., *The Possibility of Secure and Non-Secure Digital Encryption*, 1970, and archived 2014 at the Wayback Machine.

[2] Chen, Bangdao, L.H. Nguyen, and A.W. Roscoe. *When context is better than identity: authentication by context using empirical channels*. International Workshop on Security Protocols. Springer, Berlin, Heidelberg, 2011.

[3] M. Shahrzade, P. Røe, *A Survey of the State of the Art in Public Key Infrastructure*, 2003, ISBN 82-539-0502-5, Publication No. 995.

[4] J. Lopez, R. Oppliger, and G. Pernul, *Why Public Key Infrastructures have failed so far?*, Internet Research, vol. 15, pp. 544-556, 2005.

[5] L. Lamport, *Password Authentication with Insecure Communication*, Communications of the ACM, 1981, vol. 24, no 11, pp 770-772.

[6] R. Simon, Leijun Huang, E. Farrugia, *Using multiple communication channels for efficient data dissemination in wireless sensor networks*, Mobile Adhoc and Sensor Systems Conference, Washington DC, 2005.