LWIG Working Group
Internet-Draft
Intended status: Informational
Expires: December 31, 2017

C. Gomez
UPC/i2CAT
J. Crowcroft
University of Cambridge
M. Scharf
Nokia

June 29, 2017

TCP over Constrained-Node Networks
draft-gomez-lwig-tcp-constrained-node-networks-03

Abstract

This document provides a profile for the Transmission Control Protocol (TCP) over Constrained-Node Networks (CNNs). The overarching goal is to offer simple measures to allow for lightweight TCP implementation and suitable operation in such environments.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on December 31, 2017.

Table of Contents

1. Introduction 2
1.1. Conventions used in this document 3
2. Characteristics of CNNs relevant for TCP 3
3. Scenario 4
4. TCP over CNNs 4
4.1. TCP connection initiation 4
4.2. Maximum Segment Size (MSS) 5
4.3. Window Size 6
4.4. RTO estimation 6
4.5. TCP connection lifetime 7
4.5.1. Long TCP connection lifetime 7
4.5.2. Short TCP connection lifetime 7
4.6. Explicit congestion notification 8
4.7. TCP options 8
4.8. Delayed Acknowledgments 9
4.9. Explicit loss notifications 10
5. Security Considerations 10
6. Acknowledgments 10
7. Annex. TCP implementations for constrained devices 10
7.1. uIP 10
7.2. lwIP 11
7.3. RIOT 11
7.4. OpenWSN 12
7.5. TinyOS 12
7.6. Summary 12
8. References 13
8.1. Normative References 13
8.2. Informative References 15
Authors' Addresses 16

1. Introduction

The Internet Protocol suite is being used for connecting Constrained-Node Networks (CNNs) to the Internet, enabling the so-called Internet of Things (IoT) [RFC7228]. In order to meet the requirements that stem from CNNs, the IETF has produced a suite of protocols specifically designed for such environments [I-D.ietf-lwig-energy-efficient].

At the application layer, the Constrained Application Protocol (CoAP) was developed over UDP [RFC7252]. However, the integration of some CoAP deployments with existing infrastructure is being challenged by middleboxes such as firewalls, which may limit and even block UDP-

based communications. This the main reason why a CoAP over TCP specification is being developed [I-D.tschofenig-core-coap-tcp-tls].

On the other hand, other application layer protocols not specifically designed for CNNs are also being considered for the IoT space. Some examples include HTTP/2 and even HTTP/1.1, both of which run over TCP by default [RFC7540][RFC2616], and the Extensible Messaging and Presence Protocol (XMPP) [RFC 6120]. TCP is also used by non-IETF application-layer protocols in the IoT space such as MQTT and its lightweight variants [MQTTS].

This document provides a profile for TCP over CNNs. The overarching goal is to offer simple measures to allow for lightweight TCP implementation and suitable operation in such environments.

1.1. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL","SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119]

2. Characteristics of CNNs relevant for TCP

CNNs are defined in [RFC7228] as networks whose characteristics are influenced by being composed of a significant portion of constrained nodes. The latter are characterized by significant limitations on processing, memory, and energy resources, among others [RFC7228]. The first two dimensions pose constraints on the complexity and on the memory footprint of the protocols that constrained nodes can support. The latter requires techniques to save energy, such as radio duty-cycling in wireless devices [I-D.ietf-lwig-energy-efficient], as well as minimization of the number of messages transmitted/received (and their size).

Constrained nodes often use physical/link layer technologies that have been characterized as 'lossy'. Many such technologies are wireless, therefore exhibiting a relatively high bit error rate. However, some wired technologies used in the CNN space are also lossy (e.g. Power Line Communication). Transmission rates of CNN radio or wired interfaces are typically low (e.g. below 1 Mbps).

Some CNNs follow the star topology, whereby one or several hosts are linked to a central device that acts as a router connecting the CNN to the Internet. CNNs may also follow the multihop topology [RFC6606].

3. Scenario

The main scenario for use of TCP over CNNs comprises a constrained device and an unconstrained device that communicate over the Internet using TCP, possibly traversing a middlebox (e.g. a firewall, NAT, etc.). Figure 1 illustrates such scenario. Note that the scenario is asymmetric, as the unconstrained device will typically not suffer the severe constraints of the constrained device. The unconstrained device is expected to be mains-powered, to have high amount of memory and processing power, and to be connected to a resource-rich network.

Assuming that a majority of constrained devices will correspond to sensor nodes, the amount of data traffic sent by constrained devices (e.g. sensor node measurements) is expected to be higher than the amount of data traffic in the opposite direction. Nevertheless, constrained devices may receive requests (to which they may respond), commands (for configuration purposes and for constrained devices including actuators) and relatively infrequent firmware/software updates.

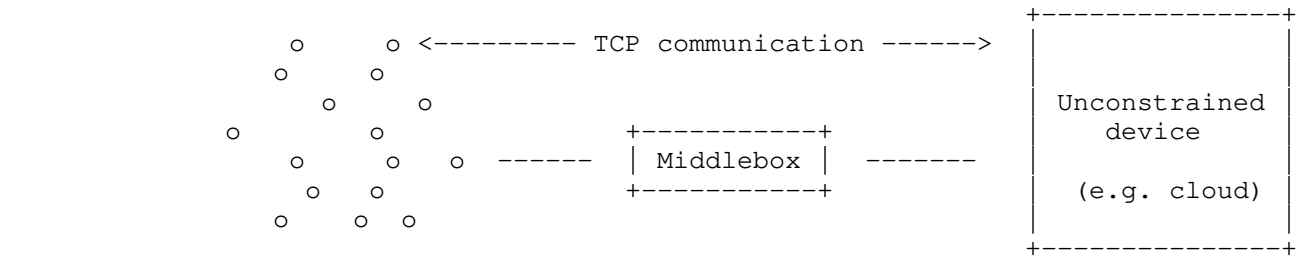

Figure 1: TCP communication between a constrained device and an unconstrained device, traversing a middlebox.

4. TCP over CNNs

4.1. TCP connection initiation

In the constrained device to unconstrained device scenario illustrated above, a TCP connection is typically initiated by the constrained device, in order for this device to support possible sleep periods to save energy.

4.2. Maximum Segment Size (MSS)

Some link layer technologies in the CNN space are characterized by a short data unit payload size, e.g. up to a few tens or hundreds of bytes. For example, the maximum frame size in IEEE 802.15.4 is 127 bytes.

6LoWPAN defined an adaptation layer to support IPv6 over IEEE 802.15.4 networks. The adaptation layer includes a fragmentation mechanism, since IPv6 requires the layer below to support an MTU of 1280 bytes [RFC2460], while IEEE 802.15.4 lacked fragmentation mechanisms. 6LoWPAN defines an IEEE 802.15.4 link MTU of 1280 bytes [RFC4944]. Other technologies, such as Bluetooth LE [RFC7668], ITU-T G.9959 [RFC7428] or DECT-ULE [RFC8105], also use 6LoWPAN-based adaptation layers in order to enable IPv6 support. These technologies do support link layer fragmentation. By exploiting this functionality, the adaptation layers that enable IPv6 over such technologies also define an MTU of 1280 bytes.

For devices using technologies with a link MTU of 1280 bytes (e.g. defined by a 6LoWPAN-based adaptation layer), in order to avoid IP layer fragmentation, the TCP MSS must not be set to a value greater than 1220 bytes in CNNs, and it must not be set to a value leading to an IPv6 datagram size exceeding 1280 bytes. (Note: IP version 6 is assumed.)

On the other hand, there exist technologies also used in the CNN space, such as Master Slave / Token Passing (TP) [RFC8163], Narrowband IoT (NB-IoT) [I-D.ietf-lpwan-overview] or IEEE 802.11ah [I-D.delcarpio-6lo-wlanah], that do not suffer the same degree of frame size limitations as the technologies mentioned above. The MTU for MS/TP is recommended to be 1500 bytes [RFC8163], the MTU in NB-IoT is 1600 bytes, and the maximum frame payload size for IEEE 802.11ah is 7991 bytes. Over such technologies, the TCP MSS may be set to a value greater than 1220 bytes, as long as IPv6 datagram size does not exceed the MTU for each technology. One consideration in this regard is that, when a node supports an MTU greater than 1280 bytes, it 'SHOULD' then support Path MTU (PMTU) discovery [RFC1981]. (Note that, as explained in RFC 1981, a minimal IPv6 implementation may 'choose to omit implementation of Path MTU Discovery'). For the sake of lightweight implementation and operation, unless applications require handling large data units (i.e. leading to an IPv6 datagram size greater than 1280 bytes), it may be desirable to limit the MTU to 1280 bytes.

4.3. Window Size

A TCP stack can reduce the implementation complexity by advertising a TCP window size of one MSS, and also transmit at most one MSS of unacknowledged data, at the cost of decreased performance. This size for receive and send window is appropriate for simple message exchanges in the CNN space, reduces implementation complexity and memory requirements, and reduces overhead (see section 4.7).

A TCP window size of one MSS follows the same rationale as the default setting for NSTART in [RFC7252], leading to equivalent operation when CoAP is used over TCP.

For devices that can afford greater TCP window size, it may be useful to allow window sizes of at least five MSSs, in order to allow Fast Retransmit and Fast Recovery [RFC5681].

4.4. RTO estimation

If a TCP sender uses very small window size and cannot use Fast Retransmit/Fast Recovery or SACK, the RTO algorithm has a larger impact on performance than for a more powerful TCP stack. In that case, RTO algorithm tuning may be considered, although careful assessment of possible drawbacks is recommended. A fundamental trade-off exists between responsiveness and correctness of RTOs [I-D.ietf-tcpm-rto-consider]. A more aggressive RTO behavior reduces wait time before retransmissions, but it also increases the probability of incurring spurious timeouts. The latter lead to unnecessary waste of potentially scarce resources in CNNs such as energy and bandwidth.

On a related note, there has been recent activity in the area of defining an adaptive RTO algorithm for CoAP (over UDP). As shown in experimental studies, the RTO estimator for CoAP defined in [I-D.ietf-core-cocoa] (hereinafter, CoCoA RTO) outperforms state-of-art algorithms designed as improvements to RFC 6298 [RFC6298] for TCP, in terms of packet delivery ratio, settling time after a burst of messages, and fairness (the latter is specially relevant in multihop networks connected to the Internet through a single device, such as a 6LoWPAN Border Router (6LBR) configured as a RPL root) [Commag]. In fact, CoCoA RTO has been designed specifically considering the challenges of CNNs, in contrast with the RFC 6298 RTO.

4.5. TCP connection lifetime

[[Note: future revisions will better separate what a TCP stack should support, or not, and how the TCP stack should be used by applications, e.g., whether to close connections or not.]]

4.5.1. Long TCP connection lifetime

In CNNs, in order to minimize message overhead, a TCP connection should be kept open as long as the two TCP endpoints have more data to exchange or it is envisaged that further segment exchanges will take place within an interval of two hours since the last segment has been sent. A greater interval may be used in scenarios where applications exchange data infrequently.

TCP keep-alive messages [RFC1122] may be supported by a server, to check whether a TCP connection is active, in order to release state of inactive connections. This may be useful for servers running on memory-constrained devices.

Since the keep-alive timer may not be set to a value lower than two hours [RFC1122], TCP keep-alive messages are not useful to guarantee that filter state records in middleboxes such as firewalls will not be deleted after an inactivity interval typically in the order of a few minutes [RFC6092]. In scenarios where such middleboxes are present, alternative measures to avoid early deletion of filter state records (which might lead to frequent establishment of new TCP connections between the two involved endpoints) include increasing the initial value for the filter state inactivity timers (if possible), and using application layer heartbeat messages.

4.5.2. Short TCP connection lifetime

A different approach to addressing the problem of traversing middleboxes that perform early filter state record deletion relies on using TCP Fast Open (TFO) [RFC7413]. In this case, instead of trying to maintain a TCP connection for long time, possibly short-lived connections can be opened between two endpoints while incurring low overhead. In fact, TFO allows data to be carried in SYN (and SYN-ACK) packets, and to be consumed immediately by the receceiving endpoint, thus reducing overhead compared with the traditional three-way handshake required to establish a TCP connection.

For security reasons, TFO requires the TCP endpoint that will open the TCP connection (which in CNNs will typically be the constrained device) to request a cookie from the other endpoint. The cookie, with a size of 4 or 16 bytes, is then included in SYN packets of subsequent connections. The cookie needs to be refreshed (and

obtained by the client) after a certain amount of time. Nevertheless, TFO is more efficient than frequently opening new TCP connections (by using the traditional three-way handshake) for transmitting new data, as long as the cookie update rate is well below the data new connection rate.

4.6. Explicit congestion notification

Explicit Congestion Notification (ECN) [RFC3168] may be used in CNNs. ECN allows a router to signal in the IP header of a packet that congestion is arising, for example when queue size reaches a certain threshold. If such a packet encapsulates a TCP data packet, an ECN-enabled TCP receiver will echo back the congestion signal to the TCP sender by setting a flag in its next TCP ACK. The sender triggers congestion control measures as if a packet loss had happened. In that case, when the congestion window of a TCP sender has a size of one segment, the TCP sender resets the retransmit timer, and will only be able to send a new packet when the retransmit timer expires [RFC3168]. Effectively, the TCP sender reduces at that moment its sending rate from 1 segment per Round Trip Time (RTT) to 1 segment per default RTO.

ECN can reduce packet losses, since congestion control measures can be applied earlier than after the reception of three duplicate ACKs (if the TCP sender window is large enough) or upon TCP sender RTO expiration [RFC2884]. Therefore, the number of retries decreases, which is particularly beneficial in CNNs, where energy and bandwidth resources are typically limited. Furthermore, latency and jitter are also reduced.

ECN is particularly appropriate in CNNs, since in these environments transactional type interactions are a dominant traffic pattern. As transactional data size decreases, the probability of detecting congestion by the presence of three duplicate ACKs decreases. In contrast, ECN can still activate congestion control measures without requiring three duplicate ACKs.

4.7. TCP options

A TCP implementation needs to support options 0, 1 and 2 [RFC793]. A TCP implementation for a constrained device that uses a single-MSS TCP receive or transmit window size may not benefit from supporting the following TCP options: Window scale [RFC1323], TCP Timestamps [RFC1323], Selective Acknowledgements (SACK) and SACK-Permitted [RFC2018]. Other TCP options should not be used, in keeping with the principle of lightweight operation.

Other TCP options should not be supported by a constrained device, in keeping with the principle of lightweight implementation and operation.

If a device, with less severe memory and processing constraints, can afford advertising a TCP window size of several MSSs, it may support the SACK option to improve performance. SACK allows a data receiver to inform the data sender of non-contiguous data blocks received, thus a sender (having previously sent the SACK-Permitted option) can avoid performing unnecessary retransmissions, saving energy and bandwidth, as well as reducing latency. The receiver supporting SACK will need to manage the reception of possible out-of-order received segments, requiring sufficient buffer space.

SACK adds 8*n+2 bytes to the TCP header, where n denotes the number of data blocks received, up to 4 blocks. For a low number of out-of-order segments, the header overhead penalty of SACK is compensated by avoiding unnecessary retransmissions.

Another potentially relevant TCP option in the context of CNNs is (TFO) [RFC7413]. As described in section 4.5.2, TFO can be used to address the problem of traversing middleboxes that perform early filter state record deletion.

4.8. Delayed Acknowledgments

A device that advertises a single-MSS receive window needs to avoid use of delayed ACKs in order to avoid contributing unnecessary delay (of up to 500 ms) to the RTT [RFC5681].

When traffic over a CNN is expected to be mostly of transactional type, with transaction size typically below one MSS, delayed ACKs are not recommended. For transactional-type traffic between a constrained device and a peer (e.g. backend infrastructure) that uses delayed ACKs, the maximum ACK rate of the peer will be typically of one ACK every 200 ms (or even lower). If in such conditions the peer device is administered by the same entity managing the constrained device, it is recommended to disable delayed ACKs at the peer side.

On the other hand, delayed ACKs allow to reduce the number of ACKs in bulk transfer type of traffic, e.g. for firmware/software updates or for transferring larger data units containing a batch of sensor readings.

4.9. Explicit loss notifications

There has been a significant body of research on solutions capable of explicitly indicating whether a TCP segment loss is due to corruption, in order to avoid activation of congestion control mechanisms [ETEN] [RFC2757]. While such solutions may provide significant improvement, they have not been widely deployed and remain as experimental work. In fact, as of today, the IETF has not standardized any such solution.

5. Security Considerations

If TFO is used, the security considerations of RFC 7413 apply.

There exist TCP options which improve TCP security. Examples include the TCP MD5 signature option [RFC2385] and the TCP Authentication Option (TCP-AO) [RFC5925]. However, both options add overhead and complexity. The TCP MD5 signature option adds 18 bytes to every segment of a connection. TCP-AO typically has a size of 16-20 bytes.

6. Acknowledgments

Carles Gomez has been funded in part by the Spanish Government (Ministerio de Educacion, Cultura y Deporte) through the Jose Castillejo grant CAS15/00336 and by European Regional Development Fund (ERDF) and the Spanish Government through project TEC2016-79988-P, AEI/FEDER, UE. Part of his contribution to this work has been carried out during his stay as a visiting scholar at the Computer Laboratory of the University of Cambridge.

The authors appreciate the feedback received for this document. The following folks provided comments that helped improve the document: Carsten Bormann, Zhen Cao, Wei Genyu, Michael Scharf, Ari Keranen, Abhijan Bhattacharyya, Andres Arcia-Moret, Yoshifumi Nishida, Joe Touch, Fred Baker, Nik Sultana, Kerry Lynn, and Erik Nordmark. Simon Brummer provided details on the RIOT TCP implementation. Xavi Vilajosana provided details on the OpenWSN TCP implementation.

7. Annex. TCP implementations for constrained devices

This section overviews the main features of TCP implementations for constrained devices.

7.1. uIP

uIP is a TCP/IP stack, targetted for 8 and 16-bit microcontrollers. uIP has been deployed with Contiki and the Arduino Ethernet shield.

A code size of ~5 kB (which comprises checksumming, IP, ICMP and TCP) has been reported for uIP [Dunk].

uIP provides a global buffer for incoming packets, of single-packet size. A buffer for outgoing data is not provided. In case of a retransmission, an application must be able to reproduce the same packet that had been transmitted.

The MSS is announced via the MSS option on connection establishment and the receive window size (of one MSS) is not modified during a connection. Stop-and-wait operation is used for sending data. Among other optimizations, this allows to avoid sliding window operations, which use 32-bit arithmetic extensively and are expensive on 8-bit CPUs.

7.2. lwIP

lwIP is a TCP/IP stack, targetted for 8- and 16-bit microcontrollers. lwIP has a total code size of ~14 kB to ~22 kB (which comprises memory management, checksumming, network interfaces, IP, ICMP and TCP), and a TCP code size of ~9 kB to ~14 kB [Dunk].

In contrast with uIP, lwIP decouples applications from the network stack. lwIP supports a TCP transmission window greater than a single segment, as well as buffering of incoming and outcoming data. Other implemented mechanisms comprise slow start, congestion avoidance, fast retransmit and fast recovery. SACK and Window Scale have been recently added to lwIP.

7.3. RIOT

The RIOT TCP implementation (called GNRC TCP) has been designed for Class 1 devices [RFC 7228]. The main target platforms are 8- and 16-bit microcontrollers. GNRC TCP offers a similar function set as uIP, but it provides and maintains an independent receive buffer for each connection. In contrast to uIP, retransmission is also handled by GNRC TCP. GNRC TCP uses a single-MSS window size, which simplifies the implementation. The application programmer does not need to know anything about the TCP internals, therefore GNRC TCP can be seen as a user-friendly uIP TCP implementation.

The MSS is set on connections establishment and cannot be changed during connection lifetime. GNRC TCP allows multiple connections in parallel, but each TCB must be allocated somewhere in the system. By default there is only enough memory allocated for a single TCP connection, but it can be increased at compile time if the user needs multiple parallel connections.

7.4. OpenWSN

The TCP implementation in OpenWSN is mostly equivalent to the uIP TCP implementation. OpenWSN TCP implementation only supports the minimum state machine functionality required. For example, it does not perform retransmissions.

7.5. TinyOS

TBD

7.6. Summary

		uIP	lwIP orig	lwIP 2.0	RIOT	OpenWSN	TinyOS
Memory	Data size	*	*	*	*	*	*
	Code size (kB)	< 5	~9 to ~14	*	*	*	*
TCP features	Window size(MSS)	1	Multiple	Multiple	1	1	*
	Slow start	No	Yes	Yes	No	No	*
	Fast rec/retx	No	Yes	Yes	No	No	*
	Keep-alive	No	*	*	No	No	*
	TFO	No	No	*	No	No	*
	ECN	No	No	*	No	No	*
	Window Scale	No	No	Yes	No	No	*
	TCP timestamps	No	No	Yes	No	No	*
	SACK	No	No	Yes	No	No	*

	Delayed ACKs	No	Yes	Yes	No	No	*

Figure 2: Summary of TCP features for differrent lightweight TCP implementations.

8. References

8.1. Normative References

[RFC1122] Braden, R., Ed., "Requirements for Internet Hosts - Communication Layers", STD 3, RFC 1122, DOI 10.17487/RFC1122, October 1989, <http://www.rfc-editor.org/info/rfc1122>.

[RFC1323] Jacobson, V., Braden, R., and D. Borman, "TCP Extensions for High Performance", RFC 1323, DOI 10.17487/RFC1323, May 1992, <http://www.rfc-editor.org/info/rfc1323>.

[RFC1981] McCann, J., Deering, S., and J. Mogul, "Path MTU Discovery for IP version 6", RFC 1981, DOI 10.17487/RFC1981, August 1996, <http://www.rfc-editor.org/info/rfc1981>.

[RFC2018] Mathis, M., Mahdavi, J., Floyd, S., and A. Romanow, "TCP Selective Acknowledgment Options", RFC 2018, DOI 10.17487/RFC2018, October 1996, <http://www.rfc-editor.org/info/rfc2018>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC2385] Heffernan, A., "Protection of BGP Sessions via the TCP MD5 Signature Option", RFC 2385, DOI 10.17487/RFC2385, August 1998, <http://www.rfc-editor.org/info/rfc2385>.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, DOI 10.17487/RFC2460, December 1998, <http://www.rfc-editor.org/info/rfc2460>.

[RFC2616] Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and T. Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.1", RFC 2616, DOI 10.17487/RFC2616, June 1999, <http://www.rfc-editor.org/info/rfc2616>.

[RFC2757] Montenegro, G., Dawkins, S., Kojo, M., Magret, V., and N. Vaidya, "Long Thin Networks", RFC 2757, DOI 10.17487/RFC2757, January 2000, <http://www.rfc-editor.org/info/rfc2757>.

[RFC2884] Hadi Salim, J. and U. Ahmed, "Performance Evaluation of Explicit Congestion Notification (ECN) in IP Networks", RFC 2884, DOI 10.17487/RFC2884, July 2000, <http://www.rfc-editor.org/info/rfc2884>.

[RFC3168] Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", RFC 3168, DOI 10.17487/RFC3168, September 2001, <http://www.rfc-editor.org/info/rfc3168>.

[RFC4944] Montenegro, G., Kushalnagar, N., Hui, J., and D. Culler, "Transmission of IPv6 Packets over IEEE 802.15.4 Networks", RFC 4944, DOI 10.17487/RFC4944, September 2007, <http://www.rfc-editor.org/info/rfc4944>.

[RFC5681] Allman, M., Paxson, V., and E. Blanton, "TCP Congestion Control", RFC 5681, DOI 10.17487/RFC5681, September 2009, <http://www.rfc-editor.org/info/rfc5681>.

[RFC5925] Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication Option", RFC 5925, DOI 10.17487/RFC5925, June 2010, <http://www.rfc-editor.org/info/rfc5925>.

[RFC6092] Woodyatt, J., Ed., "Recommended Simple Security Capabilities in Customer Premises Equipment (CPE) for Providing Residential IPv6 Internet Service", RFC 6092, DOI 10.17487/RFC6092, January 2011, <http://www.rfc-editor.org/info/rfc6092>.

[RFC6298] Paxson, V., Allman, M., Chu, J., and M. Sargent, "Computing TCP's Retransmission Timer", RFC 6298, DOI 10.17487/RFC6298, June 2011, <http://www.rfc-editor.org/info/rfc6298>.

[RFC6606] Kim, E., Kaspar, D., Gomez, C., and C. Bormann, "Problem Statement and Requirements for IPv6 over Low-Power Wireless Personal Area Network (6LoWPAN) Routing", RFC 6606, DOI 10.17487/RFC6606, May 2012, <http://www.rfc-editor.org/info/rfc6606>.

[RFC7228] Bormann, C., Ersue, M., and A. Keranen, "Terminology for Constrained-Node Networks", RFC 7228, DOI 10.17487/RFC7228, May 2014, <http://www.rfc-editor.org/info/rfc7228>.

[RFC7252] Shelby, Z., Hartke, K., and C. Bormann, "The Constrained Application Protocol (CoAP)", RFC 7252, DOI 10.17487/RFC7252, June 2014, <http://www.rfc-editor.org/info/rfc7252>.

[RFC7413] Cheng, Y., Chu, J., Radhakrishnan, S., and A. Jain, "TCP Fast Open", RFC 7413, DOI 10.17487/RFC7413, December 2014, <http://www.rfc-editor.org/info/rfc7413>.

[RFC7428] Brandt, A. and J. Buron, "Transmission of IPv6 Packets over ITU-T G.9959 Networks", RFC 7428, DOI 10.17487/RFC7428, February 2015, <http://www.rfc-editor.org/info/rfc7428>.

[RFC7540] Belshe, M., Peon, R., and M. Thomson, Ed., "Hypertext Transfer Protocol Version 2 (HTTP/2)", RFC 7540, DOI 10.17487/RFC7540, May 2015, <http://www.rfc-editor.org/info/rfc7540>.

[RFC7668] Nieminen, J., Savolainen, T., Isomaki, M., Patil, B., Shelby, Z., and C. Gomez, "IPv6 over BLUETOOTH(R) Low Energy", RFC 7668, DOI 10.17487/RFC7668, October 2015, <http://www.rfc-editor.org/info/rfc7668>.

[RFC8105] Mariager, P., Petersen, J., Ed., Shelby, Z., Van de Logt, M., and D. Barthel, "Transmission of IPv6 Packets over Digital Enhanced Cordless Telecommunications (DECT) Ultra Low Energy (ULE)", RFC 8105, DOI 10.17487/RFC8105, May 2017, <http://www.rfc-editor.org/info/rfc8105>.

[RFC8163] Lynn, K., Ed., Martocci, J., Neilson, C., and S. Donaldson, "Transmission of IPv6 over Master-Slave/Token-Passing (MS/TP) Networks", RFC 8163, DOI 10.17487/RFC8163, May 2017, <http://www.rfc-editor.org/info/rfc8163>.

8.2. Informative References

[Commag] A. Betzler, C. Gomez, I. Demirkol, J. Paradells, "CoAP Congestion Control for the Internet of Things", IEEE Communications Magazine, June 2016.

[Dunk] A. Dunkels, "Full TCP/IP for 8-Bit Architectures", 2003.

[ETEN] R. Krishnan et al, "Explicit transport error notification (ETEN) for error-prone wireless and satellite networks", Computer Networks 2004.

[I-D.delcarpio-6lo-wlanah]
Vega, L., Robles, I., and R. Morabito, "IPv6 over 802.11ah", draft-delcarpio-6lo-wlanah-01 (work in progress), October 2015.

[I-D.ietf-core-cocoa]
Bormann, C., Betzler, A., Gomez, C., and I. Demirkol, "CoAP Simple Congestion Control/Advanced", draft-ietf-core-cocoa-01 (work in progress), March 2017.

[I-D.ietf-lpwan-overview]
Farrell, S., "LPWAN Overview", draft-ietf-lpwan-overview-04 (work in progress), June 2017.

[I-D.ietf-lwig-energy-efficient]
Gomez, C., Kovatsch, M., Tian, H., and Z. Cao, "Energy-Efficient Features of Internet of Things Protocols", draft-ietf-lwig-energy-efficient-07 (work in progress), March 2017.

[I-D.ietf-tcpm-rto-consider]
Allman, M., "Retransmission Timeout Requirements", draft-ietf-tcpm-rto-consider-05 (work in progress), March 2017.

[I-D.tschofenig-core-coap-tcp-tls]
Bormann, C., Lemay, S., Technologies, Z., and H. Tschofenig, "A TCP and TLS Transport for the Constrained Application Protocol (CoAP)", draft-tschofenig-core-coap-tcp-tls-05 (work in progress), November 2015.

[MQTTS]
U. Hunkeler, H.-L. Truong, A. Stanford-Clark, "MQTT-S: A Publish/Subscribe Protocol For Wireless Sensor Networks", 2008.

Authors' Addresses

Carles Gomez
UPC/i2CAT
C/Esteve Terradas, 7
Castelldefels 08860
Spain

Email: carlesgo@entel.upc.edu

Jon Crowcroft
University of Cambridge
JJ Thomson Avenue
Cambridge, CB3 0FD
United Kingdom

Email: jon.crowcroft@cl.cam.ac.uk

Michael Scharf
Nokia
Lorenzstrasse 10
Stuttgart, 70435
Germany

Email: michael.scharf@nokia.com

Light-Weight Implementation Guidance
Internet-Draft
Intended status: Informational
Expires: May 4, 2017

M. Sethi
J. Arkko
A. Keranen
Ericsson
H. Back
Comptel

October 31, 2016

Practical Considerations and Implementation Experiences in Securing Smart Object Networks
draft-ietf-lwig-crypto-sensors-01

Abstract

This memo describes challenges associated with securing smart object devices in constrained implementations and environments. The memo describes a possible deployment model suitable for these environments, discusses the availability of cryptographic libraries for small devices, presents some preliminary experiences in implementing small devices using those libraries, and discusses trade-offs involving different types of approaches.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on May 4, 2017.

Table of Contents

1. Introduction 2
2. Related Work 3
3. Challenges 4
4. Proposed Deployment Model 5
5. Provisioning 6
6. Protocol Architecture 8
7. Code Availability 9
8. Implementation Experiences 11
9. Example Application 17
10. Design Trade-Offs 21
11. Feasibility 21
12. Freshness 22
13. Layering 24
14. Symmetric vs. Asymmetric Crypto 26
15. Security Considerations 26
16. IANA Considerations 26
17. Informative references 26
Appendix A. Acknowledgments 32
Authors' Addresses 32

1. Introduction

This memo describes challenges associated with securing smart object devices in constrained implementations and environments. In Section 3) we specifically discuss three challenges: the implementation difficulties encountered on resource-constrained platforms, the problem of provisioning keys and making the choice of implementing security at the appropriate layer.

Secondly, Section 4 discusses a deployment model that the authors are considering for constrained environments. The model requires minimal amount of configuration, and we believe it is a natural fit with the typical communication practices in smart object networking environments.

Thirdly, Section 7 discusses the availability of cryptographic libraries. Section 8 presents some experiences in implementing cryptography on small devices using those libraries, including

information about achievable code sizes and speeds on typical hardware.

Finally, Section 10 discusses trade-offs involving different types of security approaches.

2. Related Work

Constrained Application Protocol (CoAP) [RFC7252] is a light-weight protocol designed to be used in machine-to-machine applications such as smart energy and building automation. Our discussion uses this protocol as an example, but the conclusions may apply to other similar protocols. CoAP base specification [RFC7252] outlines how to use DTLS [RFC6347] and IPsec [RFC4303] for securing the protocol. DTLS can be applied with pairwise shared keys, raw public keys or with certificates. The security model in all cases is mutual authentication, so while there is some commonality to HTTP in verifying the server identity, in practice the models are quite different. The CoAP specification says little about how DTLS keys are managed. The use of IPsec with CoAP is described with regards to the protocol requirements, noting that small implementations of IKEv2 exist [RFC7815]. However, the CoAP specification is silent on policy and other aspects that are normally necessary in order to implement interoperable use of IPsec in any environment [RFC5406].

[RFC6574] gives an overview of the security discussions at the March 2011 IAB workshop on smart objects. The workshop recommended that additional work is needed in developing suitable credential management mechanisms (perhaps something similar to the Bluetooth pairing mechanism), understanding the implementability of standard security mechanisms in small devices and additional research in the area of lightweight cryptographic primitives.

[I-D.moskowitz-hip-dex] defines a light-weight version of the HIP protocol for low-power nodes. This version uses a fixed set of algorithms, Elliptic Curve Cryptography (ECC), and eliminates hash functions. The protocol still operates based on host identities, and runs end-to-end between hosts, protecting IP layer communications. [RFC6078] describes an extension of HIP that can be used to send upper layer protocol messages without running the usual HIP base exchange at all.

[I-D.daniel-6lowpan-security-analysis] makes a comprehensive analysis of security issues related to 6LoWPAN networks, but its findings also apply more generally for all low-powered networks. Some of the issues this document discusses include the need to minimize the number of transmitted bits and simplify implementations, threats in the smart object networking environments, and the suitability of

6LoWPAN security mechanisms, IPsec, and key management protocols for implementation in these environments.

[I-D.irtf-t2trg-iot-seccons] discusses the overall security problem for Internet of Things devices. It also discusses various solutions, including IKEv2/IPsec [RFC7296], TLS/SSL [RFC5246], DTLS [RFC6347], HIP [RFC7401] [I-D.moskowitz-hip-dex], PANA [RFC5191], and EAP [RFC3748]. The draft also discusses various operational scenarios, bootstrapping mechanisms, and challenges associated with implementing security mechanisms in these environments.

3. Challenges

This section discusses three challenges: implementation difficulties, practical provisioning problems, and layering and communication models.

The most often discussed issues in the security for the Internet of Things relate to implementation difficulties. The desire to build small, battery-operated, and inexpensive devices drives the creation of devices with a limited protocol and application suite. Some of the typical limitations include running CoAP instead of HTTP, limited support for security mechanisms, limited processing power for long key lengths, sleep schedule that does not allow communication at all times, and so on. In addition, the devices typically have very limited support for configuration, making it hard to set up secrets and trust anchors.

The implementation difficulties are important, but they should not be overemphasized. It is important to select the right security mechanisms and avoid duplicated or unnecessary functionality. But at the end of the day, if strong cryptographic security is needed, the implementations have to support that. Also, the use of the most lightweight algorithms and cryptographic primitives is useful, but should not be the only consideration in the design. Interoperability is also important, and often other parts of the system, such as key management protocols or certificate formats are heavier to implement than the algorithms themselves.

The second challenge relates to practical provisioning problems. These are perhaps the most fundamental and difficult issue, and unfortunately often neglected in the design. There are several problems in the provisioning and management of smart object networks:

- Small devices have no natural user interface for configuration that would be required for the installation of shared secrets and other security-related parameters. Typically, there is no keyboard, no display, and there may not even be buttons to press.

Some devices may only have one interface, the interface to the network.

- Manual configuration is rarely, if at all, possible, as the necessary skills are missing in typical installation environments (such as in family homes).

- There may be a large number of devices. Configuration tasks that may be acceptable when performed for one device may become unacceptable with dozens or hundreds of devices.

- Network configurations evolve over the lifetime of the devices, as additional devices are introduced or addresses change. Various central nodes may also receive more frequent updates than individual devices such as sensors embedded in building materials.

Finally, layering and communication models present difficulties for straightforward use of the most obvious security mechanisms. Smart object networks typically pass information through multiple participating nodes [I-D.arkko-core-sleepy-sensors] and end-to-end security for IP or transport layers may not fit such communication models very well. The primary reasons for needing middleboxes relates to the need to accommodate for sleeping nodes as well to enable the implementation of nodes that store or aggregate information.

4. Proposed Deployment Model

[I-D.arkko-core-security-arch] recognizes the provisioning model as the driver of what kind of security architecture is useful. This section re-introduces this model briefly here in order to facilitate the discussion of the various design alternatives later.

The basis of the proposed architecture are self-generated secure identities, similar to Cryptographically Generated Addresses (CGAs) [RFC3972] or Host Identity Tags (HITs) [RFC7401]. That is, we assume the following holds:

$$I = h(P|O)$$

where I is the secure identity of the device, h is a hash function, P is the public key from a key pair generated by the device, and O is optional other information. | here denotes the concatenation operator.

5. Provisioning

As it is difficult to provision security credentials, shared secrets, and policy information, the provisioning model is based only on the secure identities. A typical network installation involves physical placement of a number of devices while noting the identities of these devices. This list of short identifiers can then be fed to a central server as a list of authorized devices. Secure communications can then commence with the devices, at least as far as information from from the devices to the server is concerned, which is what is needed for sensor networks.

The above architecture is a perfect fit for sensor networks where information flows from large number of devices to small number of servers. But it is not sufficient alone for other types of applications. For instance, in actuator applications a large number of devices need to take commands from somewhere else. In such applications it is necessary to secure that the commands come from an authorized source. This can be supported, with some additional provisioning effort and optional pairing protocols. The basic provisioning approach is as described earlier, but in addition there must be something that informs the devices of the identity of the trusted server(s). There are multiple ways to provide this information. One simple approach is to feed the identities of the trusted server(s) to devices at installation time. This requires either a separate user interface, local connection (such as USB), or using the network interface of the device for configuration. In any case, as with sensor networks the amount of configuration information is minimized: just one short identity value needs to be fed in. Not both an identity and a certificate. Not shared secrets that must be kept confidential. An even simpler provisioning approach is that the devices in the device group trust each other. Then no configuration is needed at installation time. When both peers know the expected cryptographic identity of the other peer off-line, secure communications can commence. Alternatively, various pairing schemes can be employed. Note that these schemes can benefit from the already secure identifiers on the device side. For instance, the server can send a pairing message to each device after their initial power-on and before they have been paired with anyone, encrypted with the public key of the device. As with all pairing schemes that do not employ a shared secret or the secure dentity of both parties, there are some remaining vulnerabilities that may or may not be acceptable for the application in question. In any case, the secure identities help again in ensuring that the operations are as simple as possible. Only identities need to be communicated to the devices, not certificates, not shared secrets or IPsec policy rules.

Where necessary, the information collected at installation time may also include other parameters relevant to the application, such as the location or purpose of the devices. This would enable the server to know, for instance, that a particular device is the temperature sensor for the kitchen.

Collecting the identity information at installation time can be arranged in a number of ways. The authors have employed a simple but not completely secure method where the last few digits of the identity are printed on a tiny device just a few millimeters across. Alternatively, the packaging for the device may include the full identity (typically 32 hex digits), retrieved from the device at manufacturing time. This identity can be read, for instance, by a bar code reader carried by the installation personnel. (Note that the identities are not secret, the security of the system is not dependent on the identity information leaking to others. The real owner of an identity can always prove its ownership with the private key which never leaves the device.) Finally, the device may use its wired network interface or proximity-based communications, such as Near-Field Communications (NFC) or Radio-Frequency Identity tags (RFIDs). Such interfaces allow secure communication of the device identity to an information gathering device at installation time.

No matter what the method of information collection is, this provisioning model minimizes the effort required to set up the security. Each device generates its own identity in a random, secure key generation process. The identities are self-securing in the sense that if you know the identity of the peer you want to communicate with, messages from the peer can be signed by the peer's private key and it is trivial to verify that the message came from the expected peer. There is no need to configure an identity and certificate of that identity separately. There is no need to configure a group secret or a shared secret. There is no need to configure a trust anchor. In addition, the identities are typically collected anyway for application purposes (such as identifying which sensor is in which room). Under most circumstances there is actually no additional configuration effort from provisioning security.

Groups of devices can be managed through single identifiers as well. In these deployment cases it is also possible to configure the identity of an entire group of devices, rather than registering the individual devices. For instance, many installations employ a kit of devices bought from the same manufacturer in one package. It is easy to provide an identity for such a set of devices as follows:

Idev = h(Pdev|Potherdev1|Potherdev2|...|Potherdevn)

Igrp = h(Pdev1|Pdev2|...|Pdevm)

where Idev is the identity of an individual device, Pdev is the public key of that device, and Potherdevi are the public keys of other devices in the group. Now, we can define the secure identity of the group (Igrp) as a hash of all the public keys of the devices in the group (Pdevi).

The installation personnel can scan the identity of the group from the box that the kit came in, and this identity can be stored in a server that is expected to receive information from the nodes. Later when the individual devices contact this server, they will be able to show that they are part of the group, as they can reveal their own public key and the public keys of the other devices. Devices that do not belong to the kit can not claim to be in the group, because the group identity would change if any new keys were added to Igrp.

6. Protocol Architecture

As noted above, the starting point of the architecture is that nodes self-generate secure identities which are then communicated out-of-band to the peers that need to know what devices to trust. To support this model in a protocol architecture, we also need to use these secure identities to implement secure messaging between the peers, explain how the system can respond to different types of attacks such as replay attempts, and decide at what protocol layer and endpoints the architecture should use.

The deployment itself is suitable for a variety of design choices regarding layering and protocol mechanisms. [I-D.arkko-core-security-arch] was mostly focused on employing end-to-end data object security as opposed to hop-by-hop security. But other approaches are possible. For instance, HIP in its opportunistic mode could be used to implement largely the same functionality at the IP layer. However, it is our belief that the right layer for this solution is at the application layer. More specifically, in the data formats transported in the payload part of CoAP. This approach provides the following benefits:

- Ability for intermediaries to act as caches to support different sleep schedules, without the security model being impacted.

- Ability for intermediaries to be built to perform aggregation, filtering, storage and other actions, again without impacting the security of the data being transmitted or stored.

- Ability to operate in the presence of traditional middleboxes, such as a protocol translators or even NATs (not that we recommend their use in these environments).

However, as we will see later there are also some technical implications, namely that link, network, and transport layer solutions are more likely to be able to benefit from sessions where the cost of expensive operations can be amortized over multiple data transmissions. While this is not impossible in data object security solutions either, it is not the typical arrangement either.

7. Code Availability

For implementing public key cryptography on resource constrained environments, we chose Arduino Uno board [arduino-uno] as the test platform. Arduino Uno has an ATmega328 microcontroller, an 8-bit processor with a clock speed of 16 MHz, 2 kB of SRAM, and 32 kB of flash memory.

For selecting potential asymmetric cryptographic libraries, we did an extensive survey and came up with a set of possible code sources, and performed an initial analysis of how well they fit the Arduino environment. Note that the results are preliminary, and could easily be affected in any direction by implementation bugs, configuration errors, and other mistakes. Please verify the numbers before relying on them for building something. No significant effort was done to optimize ROM memory usage beyond what the libraries provided themselves, so those numbers should be taken as upper limits.

Here is the set of libraries we found:

- AvrCryptolib [avr-cryptolib]: This library provides a variety of different symmetric key algorithms such as AES and RSA as an asymmetric key algorithm. We stripped down the library to use only the required RSA components and used a separate SHA-256 implementation from the original AvrCrypto-Lib library [avr-crypto-lib]. Parts of SHA-256 and RSA algorithm implementations were written in AVR-8 bit assembly language to reduce the size and optimize the performance. The library also takes advantage of the fact that Arduino boards allow the programmer to directly address the flash memory to access constant data which can save the amount of SRAM used during execution.

- Relic-Toolkit [relic-toolkit]: This library is written entirely in C and provides a highly flexible and customizable implementation of a large variety of cryptographic algorithms. This not only includes RSA and ECC, but also pairing based asymmetric cryptography, Boneh-Lynn-Schacham, Boneh-Boyen short signatures and many more. The toolkit provides an option to build only the desired components for the required platform. While building the library, it is possible to select a variety mathematical optimizations that can be combined to obtain the desired

performance (as a general thumb rule, faster implementations require more SRAM and flash). It includes a multi precision integer math module which can be customized to use different bit-length words.

- TinyECC [tinyecc]: TinyECC was designed for using Elliptic Curve based public key cryptography on sensor networks. It is written in nesC programming language and as such is designed for specific use on TinyOS. However, the library can be ported to standard C99 either with hacked tool-chains or manually rewriting parts of the code. This allows for the library to be used on platforms that do not have TinyOS running on them. The library includes a wide variety of mathematical optimizations such as sliding window, Barrett reduction for verification, precomputation, etc. It also has one of the smallest memory footprints among the set of Elliptic Curve libraries surveyed so far. However, an advantage of Relic over TinyECC is that it can do curves over binary fields in addition to prime fields.

- Wiselib [wiselib]: Wiselib is a generic library written for sensor networks containing a wide variety of algorithms. While the stable version contains algorithms for routing only, the test version includes many more algorithms including algorithms for cryptography, localization , topology management and many more. The library was designed with the idea of making it easy to interface the library with operating systems like iSense and Contiki. However, since the library is written entirely in C++ with a template based model similar to Boost/CGAL, it can be used on any platform directly without using any of the operating system interfaces provided. This approach was taken by the authors to test the code on Arduino Uno. The structure of the code is similar to TinyECC and like TinyECC it implements elliptic curves over prime fields only. In order to make the code platform independent, no assembly level optimizations were incorporated. Since efficiency was not an important goal for the authors of the library while designing, many well known theoretical performance enhancement features were also not incorporated. Like the relic-toolkit, Wiselib is also Lesser GPL licensed.

- MatrixSSL [matrix-ssl]: This library provides a low footprint implementation of several cryptographic algorithms including RSA and ECC (with a commercial license). However, the library in the original form takes about 50 kB of ROM which is not suitable for our hardware requirements. Moreover, it is intended for 32-bit systems and the API includes functions for SSL communication rather than just signing data with private keys.

This is by no ways an exhaustive list and there exist other cryptographic libraries targeting resource-constrained devices.

8. Implementation Experiences

While evaluating the implementation experiences, we were particularly interested in the signature generation operation. This was because our example application discussed in Section 9 required only the signature generation operation on the resource-constrained platforms. We have summarized the initial results of RSA private key performance using AvrCryptolib in Table 1. All results are from a single run since repeating the test did not change (or had only minimal impact on) the results. The keys were generated separately and were hard coded into the program. All keys were generated with the value of the public exponent as 3. The performance of signing with private key was faster for smaller key lengths as was expected. However the increase in the execution time was considerable when the key size was 2048 bits. It is important to note that two different sets of experiments were performed for each key length. In the first case, the keys were loaded into the SRAM from the ROM (flash) before they were used by any of the functions. However, in the second case, the keys were addressed directly in the ROM. As was expected, the second case used less SRAM but lead to longer execution time.

More importantly, any RSA key size less than 2,048-bit should be considered legacy and insecure. The performance measurements for these keys are provided here for reference only.

Key length (bits)	Execution time (ms); key in SRAM	Memory footprint (bytes); key in SRAM	Execution time (ms); key in ROM	Memory footprint (bytes); key in ROM
64	64	40	69	32
128	434	80	460	64
512	25,076	320	27348	256
1,024	199688	640	218367	512
2,048	1587567	1,280	1740258	1,024

RSA private key operation performance

Table 1

The code size was less than 3.6 kB for all the test cases with scope for further reduction. It is also worth noting that the implementation performs basic exponentiation and multiplication

operations without using any mathematical optimizations such as Montgomery multiplication, optimized squaring, etc. as described in [rsa-high-speed]. With more SRAM, we believe that 1024/2048-bit operations can be performed in much less time as has been shown in [rsa-8bit]. 2048-bit RSA is nonetheless possible with about 1 kB of SRAM as is seen in Table 1.

In Table 2 we present the results obtained by manually porting TinyECC into C99 standard and running ECDSA signature algorithm on the Arduino Uno board. TinyECC supports a variety of SEC 2 recommended Elliptic Curve domain parameters. The execution time and memory footprint are shown next to each of the curve parameters. SHA-1 hashing algorithm included in the library was used in each of the cases. The measurements reflect the performance of elliptic curve signing only and not the SHA-1 hashing algorithm. SHA-1 is now known to be insecure and should not be used in real deployments. It is clearly observable that for similar security levels, Elliptic Curve public key cryptography outperforms RSA. These results were obtained by turning on all the optimizations. These optimizations include - Curve Specific Optimizations for modular reduction (NIST and SEC 2 field primes were chosen as pseudo-Mersenne primes), Sliding Window for faster scalar multiplication, Hybrid squaring procedure written in assembly and Weighted projective Coordinate system for efficient scalar point addition, doubling and multiplication. We did not use optimizations like Shamir Trick and Sliding Window as they are only useful for signature verification and tend to slow down the signature generation by precomputing values (we were only interested in fast signature generation). There is still some scope for optimization as not all the assembly code provided with the library could be ported to Arduino directly. Re-writing these procedures in compatible assembly would further enhance the performance.

Curve parameters	Execution time (ms)	Memory Footprint (bytes)	Comparable RSA key length
128r1	1858	776	704
128r2	2002	776	704
160k1	2228	892	1024
160r1	2250	892	1024
160r2	2467	892	1024
192k1	3425	1008	1536
192r1	3578	1008	1536

ECDSA signature performance with TinyECC

Table 2

We also performed experiments by removing the assembly code for hybrid multiplication and squaring thus using a C only form of the library. This gives us an idea of the performance that can be achieved with TinyECC on any platform regardless of what kind of OS and assembly instruction set available. The memory footprint remains the same with our without assembly code. The tables contain the maximum RAM that is used when all the possible optimizations are on. If however, the amount of RAM available is smaller in size, some of the optimizations can be turned off to reduce the memory consumption accordingly.

Curve parameters	Execution time (ms)	Memory Footprint (bytes)	Comparable RSA key length
128r1	2741	776	704
128r2	3086	776	704
160k1	3795	892	1024
160r1	3841	892	1024
160r2	4118	892	1024
192k1	6091	1008	1536
192r1	6217	1008	1536

ECDSA signature performance with TinyECC (No assembly optimizations)

Table 3

Table 4 documents the performance of Wiselib. Since there were no optimizations that could be turned on or off, we have only one set of results. By default Wiselib only supports some of the standard SEC 2 Elliptic curves. But it is easy to change the domain parameters and obtain results for for all the 128, 160 and 192-bit SEC 2 Elliptic curves. SHA-1 algorithm provided in the library was used. The measurements reflect the performance of elliptic curve signing only and not the SHA-1 hashing algorithm. SHA-1 is now known to be insecure and should not be used in real deployments. The ROM size for all the experiments was less than 16 kB.

Curve parameters	Execution time (ms)	Memory Footprint (bytes)	Comparable RSA key length
128r1	5615	732	704
128r2	5615	732	704
160k1	10957	842	1024
160r1	10972	842	1024
160r2	10971	842	1024
192k1	18814	952	1536
192r1	18825	952	1536

ECDSA signature performance with Wiselib

Table 4

For testing the relic-toolkit we used a different board because it required more RAM/ROM and we were unable to perform experiments with it on Arduino Uno. We decided to use the Arduino Mega which has the same 8-bit architecture like the Arduino Uno but has a much larger RAM/ROM for testing relic-toolkit. Again, SHA-1 hashing algorithm included in the library was used in each of the cases. The measurements reflect the performance of elliptic curve signing only and not the SHA-1 hashing algorithm. SHA-1 is now known to be insecure and should not be used in real deployments. The library does provide several alternatives with such as SHA-256.

The relic-toolkit supports Koblitz curves over prime as well as binary fields. We have experimented with Koblitz curves over binary fields only. We do not run our experiments with all the curves available in the library since the aim of this work is not prove which curves perform the fastest, and rather show that asymmetric crypto is possible on resource-constrained devices.

The results from relic-toolkit are documented in two separate tables shown in Table 5 and Table 6. The first set of results were performed with the library configured for high speed performance with no consideration given to the amount of memory used. For the second set, the library was configured for low memory usage irrespective of the execution time required by different curves. By turning on/off optimizations included in the library, a trade-off between memory and execution time between these values can be achieved.

Curve parameters	Execution time (ms)	Memory Footprint (bytes)	Comparable RSA key length
NIST K163 (assembly math)	261	2,804	1024
NIST K163	932	2750	1024
NIST B163	2243	2444	1024
NIST K233	1736	3675	2048
NIST B233	4471	3261	2048

ECDSA signature performance with relic-toolkit (Fast)

Table 5

Curve parameters	Execution time (ms)	Memory Footprint (bytes)	Comparable RSA key length
NIST K163 (assembly math)	592	2087	1024
NIST K163	2950	2215	1024
NIST B163	3213	2071	1024
NIST K233	6450	2935	2048
NIST B233	6100	2737	2048

ECDSA signature performance with relic-toolkit (Low Memory)

Table 6

It is important to note the following points about the elliptic curve measurements:

- As with the RSA measurements, curves giving less that 112-bit security are insecure and considered as legacy. The measurements are only provided for reference.

- The arduino board only provides pseudo random numbers with the random() function call. In order to create private keys with a better quality of random number, we can use a true random number generator like the one provided by TrueRandom library [truerandom], or create the keys separately on a system with a true random number generator and then use them directly in the code.

- For measuring the memory footprint of all the ECC libraries, we used the Avrora simulator [avrora]. Only stack memory was used to easily track the RAM consumption.

At the time of performing these measurements and study, it was unclear which exact elliptic curve(s) would be selected by the IETF community for use with resource-constrained devices. However now, [RFC7748] defines two elliptic curves over prime fields (Curve25519 and Curve448) that offer a high level of practical security for Diffie-Hellman key exchange. Correspondingly, there is ongoing work to specify elliptic curve signature schemes with Edwards-curve Digital Signature Algorithm (EdDSA). [I-D.irtf-cfrg-eddsa] specifies the recommended parameters for the edwards25519 and edwards448 curves. From these, curve25519 (for elliptic curve Diffie-Hellman key exchange) and edwards25519 (for elliptic curve digital signatures) are especially suitable for resource-constrained devices.

We found that the NaCl [nacl] and MicoNaCl [micronacl] libraries provide highly efficient implementations of Diffie-Hellman key exchange with curve25519. The results have shown that these libraries with curve25519 outperform other elliptic curves that provide similar levels of security. Hutter and Schwabe [naclavr] also show that signing of data using the curve Ed25519 from the NaCl library needs only 23,216,241 cycles on the same microcontroller that we used for our evaluations (Arduino Mega ATmega2560). This corresponds to about 1,4510 milliseconds of execution time. When compared to the results for other curves and libraries that offer similar level of security (such as NIST B233, NIST K233), this implementation far outperforms all others. As such, it is recommend that the IETF community uses these curves for protocol specification and implementations.

A summary library ROM use is shown in Table 7.

Library	ROM Footprint (Kilobytes)
AvrCryptolib	3.6
Wiselib	16
TinyECC	18
Relic-toolkit	29
NaCl Ed25519 [naclavr]	17-29

Summary of library ROM needs

Table 7

All the measurements here are only provided as an example to show that asymmetric-key cryptography (particularly, digital signatures) is possible on resource-constrained devices. These numbers by no way are the final source for measurements and some curves presented here may not be acceptable for real in-the-wild deployments anymore. For example, Mosdorf et al. [mosdorf] and Liu et al. [tinyecc] also document performance of ECDSA on similar resource-constrained devices.

9. Example Application

We developed an example application on the Arduino platform to use public key crypto mechanisms, data object security, and an easy provisioning model. Our application was originally developed to test different approaches to supporting communications to "always off" sensor nodes. These battery-operated or energy scavenging nodes do not have enough power to be stay on at all times. They wake up periodically and transmit their readings.

Such sensor nodes can be supported in various ways. [I-D.arkko-core-sleepy-sensors] was an early multicast-based approach. In the current application we have switched to using resource directories [I-D.ietf-core-resource-directory] and mirror proxies [I-D.vial-core-mirror-proxy] instead. Architecturally, the idea is that sensors can delegate a part of their role to a node in the network. Such a network node could be either a local resource or something in the Internet. In the case of CoAP mirror proxies, the network node agrees to hold the web resources on behalf of the sensor, while the sensor is asleep. The only role that the sensor has is to register itself at the mirror proxy, and periodically update the readings. All queries from the rest of the world go to the mirror proxy.

We constructed a system with four entities:

Sensor

This is an Arduino-based device that runs a CoAP mirror proxy client and Relic-toolkit. Relic takes 29 Kbytes of ROM, and the simple CoAP client roughly 3 kilobytes.

Mirror Proxy

This is a mirror proxy that holds resources on the sensor's behalf. The sensor registers itself to this node.

Resource Directory

While physically in the same node in our implementation, a resource directory is a logical function that allows sensors and mirror proxies to register resources in the directory. These resources can be queried by applications.

Application

This is a simple application that runs on a general purpose computer and can retrieve both registrations from the resource directory and most recent sensor readings from the mirror proxy.

The security of this system relies on an SSH-like approach. In Step 1, upon first boot, sensors generate keys and register themselves in the mirror proxy. Their public key is submitted along with the registration as an attribute in the CORE Link Format data [RFC6690].

In Step 2, when the sensor makes a sensor reading update to the mirror proxy it signs the message contents with a JOSE signature on the used JSON/SENML payload [RFC7515] [I-D.jennings-core-senml].

In Step 3, any other device in the network -- including the mirror proxy, resource directory and the application -- can check that the public key from the registration corresponds to the private key used to make the signature in the data update.

Note that checks can be done at any time and there is no need for the sensor and the checking node to be awake at the same time. In our implementation, the checking is done in the application node. This demonstrates how it is possible to implement end-to-end security even with the presence of assisting middleboxes.

To verify the feasibility of our architecture we developed a proof-of-concept prototype. In our prototype, the sensor was implemented using the Arduino Ethernet shield over an Arduino Mega board. Our implementation uses the standard C99 programming language on the

Arduino Mega board. In this prototype, the Mirror Proxy (MP) and the Resource Directory (RD) reside on the same physical host. A 64-bit x86 linux machine serves as the MP and the RD, while a similar but physically different 64-bit x86 linux machine serves as the client that requests data from the sensor. We chose the Relic library version 0.3.1 for our sample prototype as it can be easily compiled for different bit-length processors. Therefore, we were able to use it on the 8-bit processor of the Arduino Mega, as well as on the 64-bit processor of the x86 client. We used ECDSA to sign and verify data updates with the standard NIST-K163 curve parameters (163-bit Koblitz curve over binary field). While compiling Relic for our prototype, we used the fast configuration without any assembly optimizations.

The gateway implements the CoAP base specification in the Java programming language and extends it to add support for Mirror Proxy and Resource Directory REST interfaces. We also developed a minimalistic CoAP C-library for the Arduino sensor and for the client requesting data updates for a resource. The library has small SRAM requirements and uses stack-based allocation only. It is inter-operable with the Java implementation of CoAP running on the gateway. The location of the mirror proxy was pre-configured into the smart object sensor by hardcoding the IP address. We used an IPv4 network with public IP addresses obtained from a DHCP server.

Some important statistics of this prototype are listed in table Table 8. Our straw man analysis of the performance of this prototype is preliminary. Our intention was to demonstrate the feasibility of the entire architecture with public-key cryptography on an 8-bit microcontroller. The stated values can be improved further by a considerable amount. For example, the flash memory and SRAM consumption is relatively high because some of the Arduino libraries were used out-of-the- box and there are several functions which can be removed. Similarly we used the fast version of the Relic library in the prototype instead of the low memory version.

Flash memory consumption (for the entire prototype including Relic crypto + CoAP + Arduino UDP, Ethernet and DHCP Libraries)	51 kB
SRAM consumption (for the entire prototype including DHCP client + key generation + signing the hash of message + COAP + UDP + Ethernet)	4678 bytes
Execution time for creating the key pair + sending registration message + time spent waiting for acknowl-edgement	2030 ms
Execution time for signing the hash of message + sending update	987 ms
Signature overhead	42 bytes

Prototype Performance

Table 8

To demonstrate the efficacy of this communication model we compare it with a scenario where the smart objects do not transition into the energy saving sleep mode and directly serve temperature data to clients. As an example, we assume that in our architecture, the smart objects wake up once every minute to report the signed temperature data to the caching MP. If we calculate the energy consumption using the formula W = U * I * t (where U is the operating voltage, I is the current drawn and t is the execution time), and use the voltage and current values from the datasheets of the ATmega2560 (20mA-active mode and 5.4mA-sleep mode) and W5100 (183mA) chips used in the architecture, then in a one minute period, the Arduino board would consume 60.9 Joules of energy if it directly serves data and does not sleep. On the other hand, in our architecture it would only consume 2.6 Joules if it wakes up once a minute to update the MP with signed data. Therefore, a typical Li-ion battery that provides about 1800 milliamps per hour (mAh) at 5V would have a lifetime of 9 hours in the unsecured always-on scenario, whereas it would have a lifetime of about 8.5 days in the secured sleepy architecture presented. These lifetimes appear to be low because the Arduino board in the prototype uses Ethernet which is not energy efficient. The values presented only provide an estimate (ignoring the energy required to transition in and out of the sleep mode) and would vary depending on the hardware and MAC protocol used. Nonetheless, it is evident that

our architecture can increase the life of smart objects by allowing them to sleep and can ensure security at the same time.

10. Design Trade-Offs

This section attempts to make some early conclusions regarding trade-offs in the design space, based on deployment considerations for various mechanisms and the relative ease or difficulty of implementing them. This analysis looks at layering and the choice of symmetric vs. asymmetric cryptography.

11. Feasibility

The first question is whether using cryptographic security and asymmetric cryptography in particular is feasible at all on small devices. The numbers above give a mixed message. Clearly, an implementation of a significant cryptographic operation such as public key signing can be done in surprisingly small amount of code space. It could even be argued that our chosen prototype platform was unnecessarily restrictive in the amount of code space it allows: we chose this platform on purpose to demonstrate something that is as small and difficult as possible.

In reality, ROM memory size is probably easier to grow than other parameters in microcontrollers. A recent trend in microcontrollers is the introduction of 32-bit CPUs that are becoming cheaper and more easily available than 8-bit CPUs, in addition to being more easily programmable. In short, the authors do not expect the code size to be a significant limiting factor, both because of the small amount of code that is needed and because available memory space is growing rapidly.

The situation is less clear with regards to the amount of CPU power needed to run the algorithms. The demonstrated speeds are sufficient for many applications. For instance, a sensor that wakes up every now and then can likely spend a fraction of a second for the computation of a signature for the message that it is about to send. Or even spend multiple seconds in some cases. Most applications that use protocols such as DTLS that use public key cryptography only at the beginning of the session would also be fine with any of these execution times.

Yet, with reasonably long key sizes the execution times are in the seconds, dozens of seconds, or even longer. For some applications this is too long. Nevertheless, the authors believe that these algorithms can successfully be employed in small devices for the following reasons:

- With the right selection of algorithms and libraries, the execution times can actually be smaller. Using the Relic-toolkit with the NIST K163 algorithm (roughly equivalent to RSA at 1024 bits) at 0.3 seconds is a good example of this.

- As discussed in [wiman], in general the power requirements necessary to send or receive messages are far bigger than those needed to execute cryptographic operations. There is no good reason to choose platforms that do not provide sufficient computing power to run the necessary operations.

- Commercial libraries and the use of full potential for various optimizations will provide a better result than what we arrived at in this paper.

- Using public key cryptography only at the beginning of a session will reduce the per-packet processing times significantly.

12. Freshness

In our architecture, if implemented as described thus far, messages along with their signatures sent from the sensors to the mirror proxy can be recorded and replayed by an eavesdropper. The mirror proxy has no mechanism to distinguish previously received packets from those that are retransmitted by the sender or replayed by an eavesdropper. Therefore, it is essential for the smart objects to ensure that data updates include a freshness indicator. However, ensuring freshness on constrained devices can be non-trivial because of several reasons which include:

- Communication is mostly unidirectional to save energy.

- Internal clocks might not be accurate and may be reset several times during the operational phase of the smart object.

- Network time synchronization protocols such as Network Time Protocol (NTP) [RFC5905] are resource intensive and therefore may be undesirable in many smart object networks.

There are several different methods that can be used in our architecture for replay protection. The selection of the appropriate choice depends on the actual deployment scenario.

Including sequence numbers in signed messages can provide an effective method of replay protection. The mirror proxy should verify the sequence number of each incoming message and accept it only if it is greater than the highest previously seen sequence number. The mirror proxy drops any packet with a sequence number

that has already been received or if the received sequence number is greater than the highest previously seen sequence number by an amount larger than the preset threshold.

Sequence numbers can wrap-around at their maximum value and, therefore, it is essential to ensure that sequence numbers are sufficiently long. However, including long sequence numbers in packets can increase the network traffic originating from the sensor and can thus decrease its energy efficiency. To overcome the problem of long sequence numbers, we can use a scheme similar to that of Huang [huang], where the sender and receiver maintain and sign long sequence numbers of equal bit-lengths but they transmit only the least significant bits.

It is important for the smart object to write the sequence number into the permanent flash memory after each increment and before it is included in the message to be transmitted. This ensures that the sensor can obtain the last sequence number it had intended to send in case of a reset or a power failure. However, the sensor and the mirror proxy can still end up in a discordant state where the sequence number received by the mirror proxy exceeds the expected sequence number by an amount greater than the preset threshold. This may happen because of a prolonged network outage or if the mirror proxy experiences a power failure for some reason. Therefore it is essential for sensors that normally send Non-Confirmable data updates to send some Confirmable updates and re-synchronize with the mirror proxy if a reset message is received. The sensors re-synchronize by sending a new registration message with the current sequence number.

Although sequence numbers protect the system from replay attacks, a mirror proxy has no mechanism to determine the time at which updates were created by the sensor. Moreover, if sequence numbers are the only freshness indicator used, a malicious eavesdropper can induce inordinate delays to the communication of signed updates by buffering messages. It may be important in certain smart object networks for sensors to send data updates which include timestamps to allow the mirror proxy to determine the time when the update was created. For example, when the mirror proxy is collecting temperature data, it may be necessary to know when exactly the temperature measurement was made by the sensor. A simple solution to this problem is for the mirror proxy to assume that the data object was created when it receives the update. In a relatively reliable network with low RTT, it can be acceptable to make such an assumption. However most networks are susceptible to packet loss and hostile attacks making this assumption unsustainable.

Depending on the hardware used by the smart objects, they may have access to accurate hardware clocks which can be used to include

timestamps in the signed updates. These timestamps are included in addition to sequence numbers. The clock time in the smart objects can be set by the manufacturer or the current time can be communicated by the mirror proxy during the registration phase. However, these approaches require the smart objects to either rely on the long-term accuracy of the clock set by the manufacturer or to trust the mirror proxy thereby increasing the potential vulnerability of the system. The smart objects could also obtain the current time from NTP, but this may consume additional energy and give rise to security issues discussed in [RFC5905]. The smart objects could also have access to a GSM network or the Global Positioning System (GPS), and they can be used obtain the current time. Finally, if the sensors need to co-ordinate their sleep cycles, or if the mirror proxy computes an average or mean of updates collected from multiple smart objects, it is important for the network nodes to synchronize the time among them. This can be done by using existing synchronization schemes.

13. Layering

It would be useful to select just one layer where security is provided at. Otherwise a simple device needs to implement multiple security mechanisms. While some code can probably be shared across such implementations (like algorithms), it is likely that most of the code involving the actual protocol machinery cannot. Looking at the different layers, here are the choices and their implications:

link layer

This is probably the most common solution today. The biggest benefits of this choice of layer are that security services are commonly available (WLAN secrets, cellular SIM cards, etc.) and that their application protects the entire communications.

The main drawback is that there is no security beyond the first hop. This can be problematic, e.g., in many devices that communicate to a server in the Internet. A Withings scale [Withings], for instance, can support WLAN security but without some level of end-to-end security, it would be difficult to prevent fraudulent data submissions to the servers.

Another drawback is that some commonly implemented link layer security designs use group secrets. This allows any device within the local network (e.g., an infected laptop) to attack the communications.

network layer

There are a number of solutions in this space, and many new ones and variations thereof being proposed: IPsec, PANA, and so on. In general, these solutions have similar characteristics to those in the transport layer: they work across forwarding hops but only as far as to the next middlebox or application entity. There is plenty of existing solutions and designs.

Experience has shown that it is difficult to control IP layer entities from an application process. While this is theoretically easy, in practice the necessary APIs do not exist. For instance, most IPsec software has been built for the VPN use case, and is difficult or impossible to tweak to be used on a per-application basis. As a result, the authors are not particularly enthusiastic about recommending these solutions.

transport and application layer

This is another popular solution along with link layer designs. TLS with HTTP (HTTPS) and DTLS with CoAP are examples of solutions in this space, and have been proven to work well. These solutions are typically easy to take into use in an application, without assuming anything from the underlying OS, and they are easy to control as needed by the applications. The main drawback is that generally speaking, these solutions only run as far as the next application level entity. And even for this case, HTTPS can be made to work through proxies, so this limit is not unsolvable. Another drawback is that attacks on link layer, network layer and in some cases, transport layer, can not be protected against. However, if the upper layers have been protected, such attacks can at most result in a denial-of-service. Since denial-of-service can often be caused anyway, it is not clear if this is a real drawback.

data object layer

This solution does not protect any of the protocol layers, but protects individual data elements being sent. It works particularly well when there are multiple application layer entities on the path of the data. The authors believe smart object networks are likely to employ such entities for storage, filtering, aggregation and other reasons, and as such, an end-to-end solution is the only one that can protect the actual data.

The downside is that the lower layers are not protected. But again, as long as the data is protected and checked upon every time it passes through an application level entity, it is not clear that there are attacks beyond denial-of-service.

The main question mark is whether this type of a solution provides sufficient advantages over the more commonly implemented transport and application layer solutions.

14. Symmetric vs. Asymmetric Crypto

The second trade-off that is worth discussing is the use of plain asymmetric cryptographic mechanisms, plain symmetric cryptographic mechanisms, or some mixture thereof.

Contrary to popular cryptographic community beliefs, a symmetric crypto solution can be deployed in large scale. In fact, one of the largest deployment of cryptographic security, the cellular network authentication system, uses SIM cards that are based on symmetric secrets. In contrast, public key systems have yet to show ability to scale to hundreds of millions of devices, let alone billions. But the authors do not believe scaling is an important differentiator when comparing the solutions.

As can be seen from the Section 8, the time needed to calculate some of the asymmetric crypto operations with reasonable key lengths can be significant. There are two contrary observations that can be made from this. First, recent wisdom indicates that computing power on small devices is far cheaper than transmission power [wiman], and keeps on becoming more efficient very quickly. From this we can conclude that the sufficient CPU is or at least will be easily available.

But the other observation is that when there are very costly asymmetric operations, doing a key exchange followed by the use of generated symmetric keys would make sense. This model works very well for DTLS and other transport layer solutions, but works less well for data object security, particularly when the number of communicating entities is not exactly two.

15. Security Considerations

This entire memo deals with security issues.

16. IANA Considerations

There are no IANA impacts in this memo.

17. Informative references

[arduino-uno]
Arduino, "Arduino Uno", September 2015,
<http://arduino.cc/en/Main/arduinoBoardUno>.

[avr-crypto-lib]
AVR-CRYPTO-LIB, "AVR-CRYPTO-LIB", September 2015, <http://www.das-labor.org/wiki/AVR-Crypto-Lib/en>.

[avr-cryptolib]
Van der Laan, E., "AVR CRYPTOLIB", September 2015, <http://www.emsign.nl/>.

[avrora] Titzer, Ben., "Avrora", September 2015, <http://compilers.cs.ucla.edu/avrora/>.

[huang] Huang, C., "Low-overhead freshness transmission in sensor networks", 2008.

[I-D.arkko-core-security-arch]
Arkko, J. and A. Keranen, "CoAP Security Architecture", draft-arkko-core-security-arch-00 (work in progress), July 2011.

[I-D.arkko-core-sleepy-sensors]
Arkko, J., Rissanen, H., Loreto, S., Turanyi, Z., and O. Novo, "Implementing Tiny COAP Sensors", draft-arkko-core-sleepy-sensors-01 (work in progress), July 2011.

[I-D.daniel-6lowpan-security-analysis]
Park, S., Kim, K., Haddad, W., Chakrabarti, S., and J. Laganier, "IPv6 over Low Power WPAN Security Analysis", draft-daniel-6lowpan-security-analysis-05 (work in progress), March 2011.

[I-D.ietf-core-resource-directory]
Shelby, Z., Koster, M., Bormann, C., and P. Stok, "CoRE Resource Directory", draft-ietf-core-resource-directory-08 (work in progress), July 2016.

[I-D.irtf-cfrg-eddsa]
Josefsson, S. and I. Liusvaara, "Edwards-curve Digital Signature Algorithm (EdDSA)", draft-irtf-cfrg-eddsa-08 (work in progress), August 2016.

[I-D.irtf-t2trg-iot-seccons]
Garcia-Morchon, O., Kumar, S., and M. Sethi, "Security Considerations in the IP-based Internet of Things", draft-irtf-t2trg-iot-seccons-00 (work in progress), October 2016.

[I-D.jennings-core-senml]
Jennings, C., Shelby, Z., Arkko, J., and A. Keranen, "Media Types for Sensor Markup Language (SenML)", draft-jennings-core-senml-06 (work in progress), April 2016.

[I-D.moskowitz-hip-dex]
Moskowitz, R. and R. Hummen, "HIP Diet EXchange (DEX)", draft-moskowitz-hip-dex-05 (work in progress), January 2016.

[I-D.vial-core-mirror-proxy]
Vial, M., "CoRE Mirror Server", draft-vial-core-mirror-proxy-01 (work in progress), July 2012.

[matrix-ssl]
PeerSec Networks, "Matrix SSL", September 2015, <http://www.matrixssl.org/>.

[micronacl]
MicroNaCl, "The Networking and Cryptography library for microcontrollers", <http://munacl.cryptojedi.org/>.

[mosdorf] Mosdorf, M. and W. Zabolotny, "Implementation of elliptic curve cryptography for 8 bit and 32 bit embedded systems time efficiency and power consumption analysis", 2010.

[nacl] NaCl, "Networking and Cryptography library", <http://nacl.cr.yp.to/>.

[naclavr] Hutter, M. and P. Schwabe, "NaCl on 8-Bit AVR Microcontrollers", International Conference on Cryptology in Africa , Springer Berlin Heidelberg , 2013.

[relic-toolkit]
Aranha, D. and C. Gouv, "Relic Toolkit", September 2015, <http://code.google.com/p/relic-toolkit/>.

[RFC3748] Aboba, B., Blunk, L., Vollbrecht, J., Carlson, J., and H. Levkowetz, Ed., "Extensible Authentication Protocol (EAP)", RFC 3748, DOI 10.17487/RFC3748, June 2004, <http://www.rfc-editor.org/info/rfc3748>.

[RFC3972] Aura, T., "Cryptographically Generated Addresses (CGA)", RFC 3972, DOI 10.17487/RFC3972, March 2005, <http://www.rfc-editor.org/info/rfc3972>.

[RFC4303] Kent, S., "IP Encapsulating Security Payload (ESP)", RFC 4303, DOI 10.17487/RFC4303, December 2005, <http://www.rfc-editor.org/info/rfc4303>.

[RFC5191] Forsberg, D., Ohba, Y., Ed., Patil, B., Tschofenig, H., and A. Yegin, "Protocol for Carrying Authentication for Network Access (PANA)", RFC 5191, DOI 10.17487/RFC5191, May 2008, <http://www.rfc-editor.org/info/rfc5191>.

[RFC5246] Dierks, T. and E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.2", RFC 5246, DOI 10.17487/RFC5246, August 2008, <http://www.rfc-editor.org/info/rfc5246>.

[RFC5406] Bellovin, S., "Guidelines for Specifying the Use of IPsec Version 2", BCP 146, RFC 5406, DOI 10.17487/RFC5406, February 2009, <http://www.rfc-editor.org/info/rfc5406>.

[RFC5905] Mills, D., Martin, J., Ed., Burbank, J., and W. Kasch, "Network Time Protocol Version 4: Protocol and Algorithms Specification", RFC 5905, DOI 10.17487/RFC5905, June 2010, <http://www.rfc-editor.org/info/rfc5905>.

[RFC6078] Camarillo, G. and J. Melen, "Host Identity Protocol (HIP) Immediate Carriage and Conveyance of Upper-Layer Protocol Signaling (HICCUPS)", RFC 6078, DOI 10.17487/RFC6078, January 2011, <http://www.rfc-editor.org/info/rfc6078>.

[RFC6347] Rescorla, E. and N. Modadugu, "Datagram Transport Layer Security Version 1.2", RFC 6347, DOI 10.17487/RFC6347, January 2012, <http://www.rfc-editor.org/info/rfc6347>.

[RFC6574] Tschofenig, H. and J. Arkko, "Report from the Smart Object Workshop", RFC 6574, DOI 10.17487/RFC6574, April 2012, <http://www.rfc-editor.org/info/rfc6574>.

[RFC6690] Shelby, Z., "Constrained RESTful Environments (CoRE) Link Format", RFC 6690, DOI 10.17487/RFC6690, August 2012, <http://www.rfc-editor.org/info/rfc6690>.

[RFC7252] Shelby, Z., Hartke, K., and C. Bormann, "The Constrained Application Protocol (CoAP)", RFC 7252, DOI 10.17487/RFC7252, June 2014, <http://www.rfc-editor.org/info/rfc7252>.

[RFC7296] Kaufman, C., Hoffman, P., Nir, Y., Eronen, P., and T. Kivinen, "Internet Key Exchange Protocol Version 2 (IKEv2)", STD 79, RFC 7296, DOI 10.17487/RFC7296, October 2014, <http://www.rfc-editor.org/info/rfc7296>.

[RFC7401] Moskowitz, R., Ed., Heer, T., Jokela, P., and T. Henderson, "Host Identity Protocol Version 2 (HIPv2)", RFC 7401, DOI 10.17487/RFC7401, April 2015, <http://www.rfc-editor.org/info/rfc7401>.

[RFC7515] Jones, M., Bradley, J., and N. Sakimura, "JSON Web Signature (JWS)", RFC 7515, DOI 10.17487/RFC7515, May 2015, <http://www.rfc-editor.org/info/rfc7515>.

[RFC7748] Langley, A., Hamburg, M., and S. Turner, "Elliptic Curves for Security", RFC 7748, DOI 10.17487/RFC7748, January 2016, <http://www.rfc-editor.org/info/rfc7748>.

[RFC7815] Kivinen, T., "Minimal Internet Key Exchange Version 2 (IKEv2) Initiator Implementation", RFC 7815, DOI 10.17487/RFC7815, March 2016, <http://www.rfc-editor.org/info/rfc7815>.

[rsa-8bit] Gura, N., Patel, A., Wander, A., Eberle, H., and S. Shantz, "Comparing Elliptic Curve Cryptography and RSA on 8-bit CPUs", 2010.

[rsa-high-speed] Koc, C., "High-Speed RSA Implementation", November 1994, <http://cs.ucsb.edu/~koc/docs/r01.pdf>.

[tinyecc] North Carolina State University and North Carolina State University, "TinyECC", 2008, <http://discovery.csc.ncsu.edu/software/TinyECC/>.

[truerandom] Drow, C., "Truerandom", September 2015, <http://code.google.com/p/tinkerit/wiki/TrueRandom>.

[wiman] Margi, C., Oliveira, B., Sousa, G., Simplicio, M., Paulo, S., Carvalho, T., Naslund, M., and R. Gold, "Impact of Operating Systems on Wireless Sensor Networks (Security) Applications and Testbeds. In International Conference on Computer Communication Networks (ICCCN'2010) / IEEE International Workshop on Wireless Mesh and Ad Hoc Networks (WiMAN 2010), 2010, Zurich. Proceedings of ICCCN'2010/WiMAN'2010", 2010.

[wiselib] Baumgartner, T., Chatzigiannakis, I., Fekete, S., Koninis, C., Kroller, A., and A. Pyrgelis, "Wiselib", 2010, <www.wiselib.org/>.

[Withings] Withings, "The Withings scale", February 2012, <http://www.withings.com/en/bodyscale>.

Appendix A. Acknowledgments

The authors would like to thank Mats Naslund, Salvatore Loreto, Bob Moskowitz, Oscar Novo, Vlasios Tsiatsis, Daoyuan Li, Muhammad Waqas, Eric Rescorla and Tero Kivinen for interesting discussions in this problem space. The authors would also like to thank Diego Aranha for helping with the relic-toolkit configurations and Tobias Baumgartner for helping with questions regarding wiselib.

Authors' Addresses

Mohit Sethi
Ericsson
Jorvas 02420
Finland

EMail: mohit@piuha.net

Jari Arkko
Ericsson
Jorvas 02420
Finland

EMail: jari.arkko@piuha.net

Ari Keranen
Ericsson
Jorvas 02420
Finland

EMail: ari.keranen@ericsson.com

Heidi-Maria Back
Comptel
Helsinki 00181
Finland

EMail: heidi.back@comptel.com

Light-Weight Implementation Guidance
Internet-Draft
Intended status: Informational
Expires: August 30, 2018

M. Sethi
J. Arkko
A. Keranen
Ericsson
H. Back
Nokia

February 26, 2018

Practical Considerations and Implementation Experiences in Securing Smart Object Networks
draft-ietf-lwig-crypto-sensors-06

Abstract

This memo describes challenges associated with securing resource-constrained smart object devices. The memo describes a possible deployment model where resource-constrained devices sign message objects, discusses the availability of cryptographic libraries for resource-constrained devices and presents some preliminary experiences with those libraries for message signing on resource-constrained devices. Lastly, the memo discusses trade-offs involving different types of security approaches.

Table of Contents

1. Introduction 2
2. Related Work 3
3. Challenges 4
4. Proposed Deployment Model 6
4.1. Provisioning 6
4.2. Protocol Architecture 9
5. Code Availability 10
6. Implementation Experiences 11
7. Example Application 18
8. Design Trade-Offs 21
8.1. Feasibility 21
8.2. Freshness 22
8.3. Layering 24
8.4. Symmetric vs. Asymmetric Crypto 26
9. Summary 27
10. Security Considerations 27
11. IANA Considerations 27
12. Informative references 27
Appendix A. Acknowledgments 34
Authors' Addresses 34

1. Introduction

This memo describes challenges associated with securing smart object devices in constrained implementations and environments. In Section 3 we specifically discuss three challenges: the implementation difficulties encountered on resource-constrained platforms, the problem of provisioning keys and making the choice of implementing security at the appropriate layer.

Section 4 discusses a potential deployment model for constrained environments. The model requires minimal amount of configuration, and we believe it is a natural fit with the typical communication practices in smart object networking environments.

Section 5 discusses the availability of cryptographic libraries. Section 6 presents some experiences in implementing cryptography on resource-constrained devices using those libraries, including

information about achievable code sizes and speeds on typical hardware.

Finally, Section 8 discusses trade-offs involving different types of security approaches.

2. Related Work

Constrained Application Protocol (CoAP) [RFC7252] is a light-weight protocol designed to be used in machine-to-machine applications such as smart energy and building automation. Our discussion uses this protocol as an example, but the conclusions may apply to other similar protocols. The CoAP base specification [RFC7252] outlines how to use DTLS [RFC6347] and IPsec [RFC4303] for securing the protocol. DTLS can be applied with pairwise shared keys, raw public keys or with certificates. The security model in all cases is mutual authentication, so while there is some commonality to HTTP [RFC7230] in verifying the server identity, in practice the models are quite different. The use of IPsec with CoAP is described with regards to the protocol requirements, noting that lightweight implementations of IKEv2 exist [RFC7815]. However, the CoAP specification is silent on policy and other aspects that are normally necessary in order to implement interoperable use of IPsec in any environment [RFC5406].

[I-D.irtf-t2trg-iot-seccons] documents the different stages in the lifecycle of a smart object. Next, it highlights the security threats for smart objects and the challenges that one might face to protect against these threats. The document also looks at various security protocols available, including IKEv2/IPsec [RFC7296], TLS/SSL [RFC5246], DTLS [RFC6347], HIP [RFC7401], [I-D.moskowitz-hip-dex], PANA [RFC5191], and EAP [RFC3748]. Lastly, [I-D.sarikaya-t2trg-sbootstrapping] discusses bootstrapping mechanisms available for resource-constrained IoT devices.

[RFC6574] gives an overview of the security discussions at the March 2011 IAB workshop on smart objects. The workshop recommended that additional work should be undertaken in developing suitable credential management mechanisms (perhaps something similar to the Bluetooth pairing mechanism), understanding the implementability of standard security mechanisms in resource-constrained devices, and additional research in the area of lightweight cryptographic primitives.

[I-D.moskowitz-hip-dex] defines a light-weight version of the HIP protocol for low-power nodes. This version uses a fixed set of algorithms, Elliptic Curve Cryptography (ECC), and eliminates hash functions. The protocol still operates based on host identities, and runs end-to-end between hosts, protecting all IP layer

communications. [RFC6078] describes an extension of HIP that can be used to send upper layer protocol messages without running the usual HIP base exchange at all.

[I-D.daniel-6lowpan-security-analysis] makes a comprehensive analysis of security issues related to 6LoWPAN networks, but its findings also apply more generally for all low-powered networks. Some of the issues this document discusses include the need to minimize the number of transmitted bits and simplify implementations, threats in the smart object networking environments, and the suitability of 6LoWPAN security mechanisms, IPsec, and key management protocols for implementation in these environments.

3. Challenges

This section discusses three challenges: 1) implementation difficulties, 2) practical provisioning problems, 3) layering and communication models.

One of the most often discussed issues in the security for the Internet of Things relate to implementation difficulties. The desire to build resource-constrained, battery-operated, and inexpensive devices drives the creation of devices with a limited protocol and application suite. Some of the typical limitations include running CoAP instead of HTTP, limited support for security mechanisms, limited processing power for long key lengths, sleep schedule that does not allow communication at all times, and so on. In addition, the devices typically have very limited support for configuration, making it hard to set up secrets and trust anchors.

The implementation difficulties are important, but they should not be overemphasized. It is important to select the right security mechanisms and avoid duplicated or unnecessary functionality. But at the end of the day, if strong cryptographic security is needed, the implementations have to support that. It is important for developers and product designers to determine what security threats they want to tackle and the resulting security requirements before selecting the hardware. Often, development work in the wild happens in the wrong order: a particular platform with a resource-constrained microcontroller is chosen first, and then the security features that can fit on it are decided. Also, the use of the most lightweight algorithms and cryptographic primitives is useful, but should not be the only consideration in the design and development. Interoperability is also important, and often other parts of the system, such as key management protocols or certificate formats are heavier to implement than the algorithms themselves.

The second challenge relates to practical provisioning problems. This is perhaps the most fundamental and difficult issue, and unfortunately often neglected in the design. There are several problems in the provisioning and management of smart object networks:

- Resource-constrained devices have no natural user interface for configuration that would be required for the installation of shared secrets and other security-related parameters. Typically, there is no keyboard, no display, and there may not even be buttons to press. Some devices may only have one interface, the interface to the network.

- Manual configuration is rarely, if at all, possible, as the necessary skills are missing in typical installation environments (such as in family homes).

- There may be a large number of devices. Configuration tasks that may be acceptable when performed for one device may become unacceptable with dozens or hundreds of devices.

- Smart object networks may rely on different radio technologies. Provisioning methods that rely on specific link-layer features may not work with other radio technologies in a heterogeneous network.

- Network configurations evolve over the lifetime of the devices, as additional devices are introduced or addresses change. Various central nodes may also receive more frequent updates than individual devices such as sensors embedded in building materials.

In light of the above challenges, resource-constrained devices are often shipped with a single static identity. In many cases, it is a single raw public key. These long-term static identities makes it easy to track the devices (and their owners) when they move. The static identities may also allow an attacker to track these devices across ownership changes.

Finally, layering and communication models present difficulties for straightforward use of the most obvious security mechanisms. Smart object networks typically pass information through multiple participating nodes [I-D.arkko-core-sleepy-sensors] and end-to-end security for IP or transport layers may not fit such communication models very well. The primary reasons for needing middleboxes relates to the need to accommodate for sleeping nodes as well to enable the implementation of nodes that store or aggregate information.

4. Proposed Deployment Model

[I-D.arkko-core-security-arch] recognizes the provisioning model as the driver of what kind of security architecture is useful. This section re-introduces this model briefly here in order to facilitate the discussion of the various design alternatives later.

The basis of the proposed architecture are self-generated secure identities, similar to Cryptographically Generated Addresses (CGAs) [RFC3972] or Host Identity Tags (HITs) [RFC7401]. That is, we assume the following holds:

$$I = h(P|O)$$

where I is the secure identity of the device, h is a hash function, P is the public key from a key pair generated by the device, and O is optional other information. | here denotes the concatenation operator.

4.1. Provisioning

As it is difficult to provision security credentials, shared secrets, and policy information, the provisioning model is based only on the secure identities. A typical network installation involves physical placement of a number of devices while noting the identities of these devices. This list of short identifiers can then be fed to a central server as a list of authorized devices. Secure communications can then commence with the devices, at least as far as information from from the devices to the server is concerned, which is what is needed for sensor networks.

The above architecture is a perfect fit for sensor networks where information flows from large number of devices to small number of servers. But it is not sufficient alone for other types of applications. For instance, in actuator applications a large number of devices need to take commands from somewhere else. In such applications it is necessary to secure that the commands come from an authorized source.

This can be supported, with some additional provisioning effort and optional pairing protocols. The basic provisioning approach is as described earlier, but in addition there must be something that informs the devices of the identity of the trusted server(s). There are multiple ways to provide this information. One simple approach is to feed the identities of the trusted server(s) to devices at installation time. This requires either a separate user interface, local connection (such as USB), or using the network interface of the device for configuration. In any case, as with sensor networks the

amount of configuration information is minimized: just one short identity value needs to be fed in (not both an identity and certificate or shared secrets that must be kept confidential). An even simpler provisioning approach is that the devices in the device group trust each other. Then no configuration is needed at installation time.

Once both the parties interested in communicating know the expected cryptographic identity of the other off-line, secure communications can commence. Alternatively, various pairing schemes can be employed. Note that these schemes can benefit from the already secure identifiers on the device side. For instance, the server can send a pairing message to each device after their initial power-on and before they have been paired with anyone, encrypted with the public key of the device. As with all pairing schemes that do not employ a shared secret or the secure identity of both parties, there are some remaining vulnerabilities that may or may not be acceptable for the application in question. For example, many leap-of-faith or trust-on-first-use based pairing methods assume that the attacker is not present during the initial setup. Therefore, they are vulnerable to eavesdropping or man-in-the-middle (MitM) attacks.

In any case, the secure identities help again in ensuring that the operations are as simple as possible. Only identities need to be communicated to the devices, not certificates, not shared secrets or e.g. IPsec policy rules.

Where necessary, the information collected at installation time may also include other parameters relevant to the application, such as the location or purpose of the devices. This would enable the server to know, for instance, that a particular device is the temperature sensor for the kitchen.

Collecting the identity information at installation time can be arranged in a number of ways. One simple but not completely secure method where the last few digits of the identity are printed on a tiny device just a few millimeters across. Alternatively, the packaging for the device may include the full identity (typically 32 hex digits), retrieved from the device at manufacturing time. This identity can be read, for instance, by a bar code reader carried by the installation personnel. (Note that the identities are not secret, the security of the system is not dependent on the identity information leaking to others. The real owner of an identity can always prove its ownership with the private key which never leaves the device.) Finally, the device may use its wired network interface or proximity-based communications, such as Near-Field Communications (NFC) or Radio-Frequency Identity tags (RFIDs). Such interfaces

allow secure communication of the device identity to an information gathering device at installation time.

No matter what the method of information collection is, this provisioning model minimizes the effort required to set up the security. Each device generates its own identity in a random, secure key generation process. The identities are self-securing in the sense that if you know the identity of the peer you want to communicate with, messages from the peer can be signed by the peer's private key and it is trivial to verify that the message came from the expected peer. There is no need to configure an identity and certificate of that identity separately. There is no need to configure a group secret or a shared secret. There is no need to configure a trust anchor. In addition, the identities are typically collected anyway for application purposes (such as identifying which sensor is in which room). Under most circumstances there is actually no additional configuration effort from provisioning security.

As discussed in the previous section, long-term static identities negatively affect the privacy of the devices and their owners. Therefore, it is beneficial for devices to generate new identities at appropriate times during their lifecycle. For example, after a factory reset or an ownership handover. Thus, in our proposed deployment model, the devices would generate a new asymmetric key pair and use the new public-key P' to generate the new identity I'. It is also desirable that these identities are only used during the provisioning stage. Temporary identities (such as dynamic IPv6 addresses) can be used for network communication protocols once the device is operational.

Groups of devices can be managed through single identifiers as well. In these deployment cases it is also possible to configure the identity of an entire group of devices, rather than registering the individual devices. For instance, many installations employ a kit of devices bought from the same manufacturer in one package. It is easy to provide an identity for such a set of devices as follows:

Idev = h(Pdev|Potherdev1|Potherdev2|...|Potherdevn)

Igrp = h(Pdev1|Pdev2|...|Pdevm)

where Idev is the identity of an individual device, Pdev is the public key of that device, and Potherdevi are the public keys of other devices in the group, n is all the devices in the group except the device with Pdev as its public key, and m is the total number of devices in the group. Now, we can define the secure identity of the group (Igrp) as a hash of all the public keys of the devices in the group (Pdevi).

The installation personnel can scan the identity of the group from the box that the kit came in, and this identity can be stored in a server that is expected to receive information from the nodes. Later when the individual devices contact this server, they will be able to show that they are part of the group, as they can reveal their own public key and the public keys of the other devices. Devices that do not belong to the kit can not claim to be in the group, because the group identity would change if any new keys were added to the identity of the group (Igrp).

4.2. Protocol Architecture

As noted above, the starting point of the architecture is that nodes self-generate secure identities which are then communicated out-of-band to the peers that need to know what devices to trust. To support this model in a protocol architecture, we also need to use these secure identities to implement secure messaging between the peers, explain how the system can respond to different types of attacks such as replay attempts, and decide at what protocol layer and endpoints the architecture should use.

The deployment itself is suitable for a variety of design choices regarding layering and protocol mechanisms. [I-D.arkko-core-security-arch] was mostly focused on employing end-to-end data object security as opposed to hop-by-hop security. But other approaches are possible. For instance, HIP in its opportunistic mode could be used to implement largely the same functionality at the IP layer. However, it is our belief that the right layer for this solution is at the application layer. More specifically, in the data formats transported in the payload part of CoAP. This approach provides the following benefits:

- Ability for intermediaries to act as caches to support different sleep schedules, without the security model being impacted.

- Ability for intermediaries to be built to perform aggregation, filtering, storage and other actions, again without impacting the security of the data being transmitted or stored.

- Ability to operate in the presence of traditional middleboxes, such as a protocol translators or even NATs (not that we recommend their use in these environments).

However, as we will see later there are also some technical implications, namely that link, network, and transport layer solutions are more likely to be able to benefit from sessions where the cost of expensive operations can be amortized over multiple data

transmissions. While this is not impossible in data object security solutions, it is generally not the typical arrangement.

5. Code Availability

For implementing public key cryptography on resource constrained environments, we chose Arduino Uno board [arduino-uno] as the test platform. Arduino Uno has an ATmega328 microcontroller, an 8-bit processor with a clock speed of 16 MHz, 2 kB of RAM, and 32 kB of flash memory. Our choice of a 8-bit platform may seem surprising since cheaper and more energy-efficient 32-bit platforms are available. However, our intention was to evaluate the performance of public-key cryptography on the most resource-constrained platforms available. It is reasonable to expect better performance results from 32-bit microcontrollers.

For selecting potential asymmetric cryptographic libraries, we surveyed and came up with a set of possible code sources, and performed an initial analysis of how well they fit the Arduino environment. Note that the results are preliminary, and could easily be affected in any direction by implementation bugs, configuration errors, and other mistakes. It is advisable to verify the numbers before relying on them for building something. No significant effort was done to optimize ROM memory usage beyond what the libraries provided themselves, so those numbers should be taken as upper limits.

Here is the set of libraries we found:

- AvrCryptolib [avr-cryptolib]: This library provides symmetric key algorithms such as AES. It provides RSA as an asymmetric key algorithm. Parts of the library were written in AVR-8 bit assembly language to reduce the size and optimize the performance.

- Relic-Toolkit [relic-toolkit]: This library is written entirely in C and provides a highly flexible and customizable implementation of a large variety of cryptographic algorithms. This not only includes RSA and ECC, but also pairing based asymmetric cryptography, Boneh-Lynn-Schacham, Boneh-Boyen short signatures. The library has also added support for curve25519 (for elliptic curve Diffie-Hellman key exchange) [RFC7748] and edwards25519 (for elliptic curve digital signatures) [RFC8032]. The toolkit provides an option to build only the desired components for the required platform.

- TinyECC [tinyecc]: TinyECC was designed for using elliptic curve based public key cryptography on sensor networks. It is written in the nesC programming language [nesC] and as such is designed

for specific use on TinyOS. However, the library can be ported to standard C either with tool-chains or manually rewriting parts of the code. It also has one of the smallest memory footprints among the set of elliptic curve libraries surveyed so far.

- Wiselib [wiselib]: Wiselib is a generic library written for sensor networks containing a wide variety of algorithms. While the stable version contains algorithms for routing only, the test version includes many more algorithms including algorithms for cryptography, localization, topology management and many more. The library was designed with the idea of making it easy to interface the library with operating systems like iSense and Contiki. However, since the library is written entirely in C++ with a template based model similar to Boost/CGAL, it can be used on any platform directly without using any of the operating system interfaces provided. This approach was taken to test the code on Arduino Uno.

- MatrixSSL [matrix-ssl]: This library provides a low footprint implementation of several cryptographic algorithms including RSA and ECC (with a commercial license). The library in the original form takes about 50 kB of ROM and is intended for 32-bit platforms.

This is by no ways an exhaustive list and there exist other cryptographic libraries targeting resource-constrained devices.

There are also a number of operating systems that are specifically targeted for resource-constrained devices. These operating systems may included libraries and code for security. Hahm et al.[hahmos] conduct a survey of such operating systems. The ARM mbed OS [mbed] is one such operating system that provides various cryptographic primitives that are necessary for SSL/TLS protocol implementation as well as X509 certificate handling. The library provides an API for developer with a minimal code footprint. It is intended for various ARM platforms such as ARM Cortex M0, ARM Cortex M0+ and ARM Cortex M3.

6. Implementation Experiences

While evaluating the implementation experiences, we were particularly interested in the signature generation operation. This was because our example application discussed in Section 7 required only the signature generation operation on the resource-constrained platforms. We have summarized the initial results of RSA private key exponentiation performance using AvrCryptolib [avr-crypto-lib] in Table 1. All results are from a single run since repeating the test did not change (or had only minimal impact on) the results. The

execution time for a key size of 2048 bits was inordinately long and would be a deterrent in real-world deployments.

Key length (bits)	Execution time (ms); key in RAM	Memory footprint (bytes); key in RAM
2048	1587567	1280

RSA private key operation performance

Table 1

The code size was about 3.6 kB with potential for further reduction. It is also worth noting that the implementation performs basic exponentiation and multiplication operations without using any mathematical optimizations such as Montgomery multiplication, optimized squaring, etc. as described in [rsa-high-speed]. With more RAM, we believe that 2048-bit operations can be performed in much less time as has been shown in [rsa-8bit].

In Table 2 we present the results obtained by manually porting TinyECC into C99 standard and running the Elliptic Curve Digital Signature Algorithm (ECDSA) on the Arduino Uno board. TinyECC supports a variety of SEC 2 recommended Elliptic Curve domain parameters [sec2ecc]. The execution time and memory footprint are shown next to each of the curve parameters. These results were obtained by turning on all the optimizations and using assembly code where available.

The results from the performance evaluation of ECDSA in the following tables also contains a column stating the approximate comparable RSA key length as documented in [sec2ecc]. It is clearly observable that for similar security levels, Elliptic Curve public key cryptography outperforms RSA.

Curve parameters	Execution time (ms)	Memory Footprint (bytes)	Comparable RSA key length
secp160k1	2228	892	1024
secp160r1	2250	892	1024
secp160r2	2467	892	1024
secp192k1	3425	1008	1536
secp192r1	3578	1008	1536

Performance of ECDSA sign operation with TinyECC

Table 2

We also performed experiments by removing the assembly optimization and using a C only form of the library. This gives us an idea of the performance that can be achieved with TinyECC on any platform regardless of what kind of OS and assembly instruction set available. The memory footprint remains the same with or without assembly code. The tables contain the maximum RAM that is used when all the possible optimizations are on. If however, the amount of RAM available is smaller in size, some of the optimizations can be turned off to reduce the memory consumption accordingly.

Curve parameters	Execution time (ms)	Memory Footprint (bytes)	Comparable RSA key length
secp160k1	3795	892	1024
secp160r1	3841	892	1024
secp160r2	4118	892	1024
secp192k1	6091	1008	1536
secp192r1	6217	1008	1536

Performance of ECDSA sign operation with TinyECC (No assembly optimizations)

Table 3

Table 4 documents the performance of Wiselib. Since there were no optimizations that could be turned on or off, we have only one set of results. By default Wiselib only supports some of the standard SEC 2 Elliptic curves, but it is easy to change the domain parameters and

obtain results for all the 128, 160 and 192-bit SEC 2 Elliptic curves. The ROM size for all the experiments was less than 16 kB.

Curve parameters	Execution time (ms)	Memory Footprint (bytes)	Comparable RSA key length
secp160k1	10957	842	1024
secp160r1	10972	842	1024
secp160r2	10971	842	1024
secp192k1	18814	952	1536
secp192r1	18825	952	1536

Performance ECDSA sign operation with Wiselib

Table 4

For testing the relic-toolkit we used a different board because it required more RAM/ROM and we were unable to perform experiments with it on Arduino Uno. Arduino Mega has the same 8-bit architecture like the Arduino Uno but has a much larger RAM/ROM. We used Arduino Mega for experimenting with the relic-toolkit. Again, it is important to mention that we used Arduino as it is a convenient prototyping platform. Our intention was to demonstrate the feasibility of the entire architecture with public key cryptography on an 8-bit microcontroller. However it is important to state that 32-bit microcontrollers are much more easily available, at lower costs and are more power efficient. Therefore, real deployments are better off using 32-bit microcontrollers that allow developers to include the necessary cryptographic libraries. There is no good reason to choose platforms that do not provide sufficient computing power to run the necessary cryptographic operations.

The relic-toolkit supports Koblitz curves over prime as well as binary fields. We have experimented with Koblitz curves over binary fields only. We do not run our experiments with all the curves available in the library since the aim of this work is not prove which curves perform the fastest, and rather show that asymmetric cryptography is possible on resource-constrained devices.

The results from relic-toolkit are documented in two separate tables shown in Table 5 and Table 6. The first set of results were performed with the library configured for high speed performance with no consideration given to the amount of memory used. For the second set, the library was configured for low memory usage irrespective of the execution time required by different curves. By turning on/off

optimizations included in the library, a trade-off between memory and execution time between these values can be achieved.

Curve parameters	Execution time (ms)	Memory Footprint (bytes)	Comparable RSA key length
sect163k1 (assembly math)	261	2804	1024
sect163k1	932	2750	1024
sect163r2	2243	2444	1024
sect233k1	1736	3675	2048
sect233r1	4471	3261	2048

Performance of ECDSA sign operation with relic-toolkit (Fast)

Table 5

Curve parameters	Execution time (ms)	Memory Footprint (bytes)	Comparable RSA key length
sect163k1 (assembly math)	592	2087	1024
sect163k1	2950	2215	1024
sect163r2	3213	2071	1024
sect233k1	6450	2935	2048
sect233r1	6100	2737	2048

Performance of ECDSA sign operation with relic-toolkit (Low Memory)

Table 6

It is important to note the following points about the elliptic curve measurements:

- Some boards (e.g. Arduino Uno) do not provide a hardware random number generator. On such boards, obtaining cryptographic-quality randomness is a challenge. Real-world deployments must rely on a hardware random number generator for cryptographic operations such as generating a public-private key pair. The Nordic nRF52832 board [nordic] for example provides a hardware random number generator. A detailed discussion on requirements and best

practices for cryptographic-quality randomness is documented in [RFC4086]

- For measuring the memory footprint of all the ECC libraries, we used the Avrora simulator [avrora]. Only stack memory was used to easily track the RAM consumption.

Tschofenig and Pegourie-Gonnard [armecdsa] have also evaluated the performance of Elliptic Curve Cryptography (ECC) on ARM Coretex platform. The results for ECDSA sign operation shown in Table 7 are performed on a Freescale FRDM-KL25Z board [freescale] that has a ARM Cortex-M0+ 48MHz microcontroller with 128kB of flash memory and 16kB of RAM. The sliding window technique for efficient exponentiation was used with a window size of 2. All other optimizations were disabled for these measurements.

Curve parameters	Execution time (ms)	Comparable RSA key length
secp192r1	2165	1536
secp224r1	3014	2048
secp256r1	3649	2048

Performance of ECDSA sign operation with ARM mbed TLS stack on Freescale FRDM-KL25Z

Table 7

Tschofenig and Pegourie-Gonnard [armecdsa] also measured the performance of curves on a ST Nucleo F091 (STM32F091RCT6) board [stnucleo] that has a ARM Cortex-M0 48MHz microcontroller with 256 kB of flash memory and 32kB of RAM. The execution time for ECDSA sign operation with different curves is shown in Table 8. The sliding window technique for efficient exponentiation was used with a window size of 7. Fixed point optimization and NIST curve specific optimizations were used for these measurements.

Curve parameters	Execution time (ms)	Comparable RSA key length
secp192k1	291	1536
secp192r1	225	1536
secp224k1	375	2048
secp224r1	307	2048
secp256k1	486	2048
secp256r1	459	2048
secp384r1	811	7680
secp521r1	1602	15360

ECDSA signature performance with ARM mbed TLS stack on ST Nucleo F091 (STM32F091RCT6)

Table 8

Finally, Tschofenig and Pegourie-Gonnard [armecdsa] also measured the RAM consumption by calculating the heap consumed for the cryptographic operations using a custom memory allocation handler. They did not measure the minimal stack memory consumption. Depending on the curve and the different optimizations enable or disabled, the memory consumption for the ECDSA sign operation varied from 1500 bytes to 15000 bytes.

At the time of performing these measurements and study, it was unclear which exact elliptic curve(s) would be selected by the IETF community for use with resource-constrained devices. However now, [RFC7748] defines two elliptic curves over prime fields (Curve25519 and Curve448) that offer a high level of practical security for Diffie-Hellman key exchange. Correspondingly, there is ongoing work to specify elliptic curve signature schemes with Edwards-curve Digital Signature Algorithm (EdDSA). [RFC8032] specifies the recommended parameters for the edwards25519 and edwards448 curves. From these, curve25519 (for elliptic curve Diffie-Hellman key exchange) and edwards25519 (for elliptic curve digital signatures) are especially suitable for resource-constrained devices.

We found that the NaCl [nacl] and MicoNaCl [micronacl] libraries provide highly efficient implementations of Diffie-Hellman key exchange with curve25519. The results have shown that these libraries with curve25519 outperform other elliptic curves that provide similar levels of security. Hutter and Schwabe [naclavr] also show that signing of data using the curve Ed25519 from the NaCl library needs only 23216241 cycles on the same microcontroller that we used for our evaluations (Arduino Mega ATmega2560). This

corresponds to about 1451 milliseconds of execution time. When compared to the results for other curves and libraries that offer similar level of security (such as NIST B233, NIST K233), this implementation far outperforms all others. As such, it is recommend that the IETF community uses these curves for protocol specification and implementations.

A summary library flash memory use is shown in Table 9.

Library	Flash memory Footprint (Kilobytes)
AvrCryptolib	3.6
Wiselib	16
TinyECC	18
Relic-toolkit	29
NaCl Ed25519 [naclavr]	17-29

Summary of library flash memory consumption

Table 9

All the measurements here are only provided as an example to show that asymmetric-key cryptography (particularly, digital signatures) is possible on resource-constrained devices. These numbers by no way are the final source for measurements and some curves presented here may not be acceptable for real in-the-wild deployments anymore. For example, Mosdorf et al. [mosdorf] and Liu et al. [tinyecc] also document performance of ECDSA on similar resource-constrained devices.

7. Example Application

We developed an example application on the Arduino platform to use public key crypto mechanisms, data object security, and an easy provisioning model. Our application was originally developed to test different approaches to supporting communications to "always off" sensor nodes. These battery-operated or energy scavenging nodes do not have enough power to stay on at all times. They wake up periodically and transmit their readings.

Such sensor nodes can be supported in various ways. [I-D.arkko-core-sleepy-sensors] was an early multicast-based approach. In the current application we have switched to using resource directories [I-D.ietf-core-resource-directory] and publish-subscribe brokers [I-D.ietf-core-coap-pubsub] instead. Architecturally, the idea is that sensors can delegate a part of

their role to a node in the network. Such a network node could be either a local resource or something in the Internet. In the case of CoAP publish-subscribe brokers, the network node agrees to hold the web resources on behalf of the sensor, while the sensor is asleep. The only role that the sensor has is to register itself at the publish-subscribe broker, and periodically update the readings. All queries from the rest of the world go to the publish-subscribe broker.

We constructed a system with four entities:

Sensor

This is an Arduino-based device that runs a CoAP publish-subscribe broker client and Relic-toolkit. Relic takes 29 Kbytes of flash memory, and the simple CoAP client roughly 3 kilobytes.

Publish-Subscribe Broker

This is a publish-subscribe broker that holds resources on the sensor's behalf. The sensor registers itself to this node.

Resource Directory

While physically in the same node in our implementation, a resource directory is a logical function that allows sensors and publish-subscribe brokers to register resources in the directory. These resources can be queried by applications.

Application

This is a simple application that runs on a general purpose computer and can retrieve both registrations from the resource directory and most recent sensor readings from the publish-subscribe broker.

The security of this system relies on an SSH-like approach. In Step 1, upon first boot, sensors generate keys and register themselves in the publish-subscribe broker. Their public key is submitted along with the registration as an attribute in the CORE Link Format data [RFC6690].

In Step 2, when the sensor makes a measurement, it sends an update to the publish-subscribe broker and signs the message contents with a JOSE signature on the used JSON/SENML payload [RFC7515] [I-D.ietf-core-senml]. The sensor can also alternatively use CBOR Object Signing and Encryption (COSE) [RFC8152] for signing the sensor measurement.

In Step 3, any other device in the network -- including the publish-subscribe broker, resource directory and the application -- can check that the public key from the registration corresponds to the private key used to make the signature in the data update.

Note that checks can be done at any time and there is no need for the sensor and the checking node to be awake at the same time. In our implementation, the checking is done in the application node. This demonstrates how it is possible to implement end-to-end security even with the presence of assisting middleboxes.

To verify the feasibility of our architecture we developed a proof-of-concept prototype. In our prototype, the sensor was implemented using the Arduino Ethernet shield over an Arduino Mega board. Our implementation uses the standard C99 programming language on the Arduino Mega board. In this prototype, the publish-subscribe broker and the Resource Directory (RD) reside on the same physical host. A 64-bit x86 linux machine serves as the broker and the RD, while a similar but physically distinct 64-bit x86 linux machine serves as the client that requests data from the sensor. We chose the Relic library version 0.3.1 for our sample prototype as it can be easily compiled for different bit-length processors. Therefore, we were able to use it on the 8-bit processor of the Arduino Mega, as well as on the 64-bit processor of the x86 client. We used ECDSA to sign and verify data updates with the standard NIST-K163 curve parameters. While compiling Relic for our prototype, we used the fast configuration without any assembly optimizations.

The gateway implements the CoAP base specification in the Java programming language and extends it to add support for publish-subscribe broker and Resource Directory REST interfaces. We also developed a minimalistic CoAP C-library for the Arduino sensor and for the client requesting data updates for a resource. The library has small RAM requirements and uses stack-based allocation only. It is interoperable with the Java implementation of CoAP running on the gateway. The location of the resource directory was configured into the smart object sensor by hardcoding the IP address. A real implementation based on this prototype would instead use the domain name system for obtaining the location of the resource directory.

Our intention was to demonstrate that it is possible to implement the entire architecture with public-key cryptography on an 8-bit microcontroller. The stated values can be improved further by a considerable amount. For example, the flash memory and RAM consumption is relatively high because some of the Arduino libraries were used out-of-the-box and there are several functions which can be removed. Similarly we used the fast version of the Relic library in the prototype instead of the low memory version. However, it is

important to note that this was only a research prototype to verify the feasibility of this architecture and as stated elsewhere, most modern development boards have a 32-bit microcontroller since they are more economical and have better energy efficiency.

8. Design Trade-Offs

This section attempts to make some early conclusions regarding trade-offs in the design space, based on deployment considerations for various mechanisms and the relative ease or difficulty of implementing them. In particular, this analysis looks at layering, freshness and the choice of symmetric vs. asymmetric cryptography.

8.1. Feasibility

The first question is whether using cryptographic security and asymmetric cryptography in particular is feasible at all on resource-constrained devices. The numbers above give a mixed message. Clearly, an implementation of a significant cryptographic operation such as public key signing can be done in surprisingly small amount of code space. It could even be argued that our chosen prototype platform was unnecessarily restrictive in the amount of code space it allows: we chose this platform on purpose to demonstrate something that is as resource-constrained and difficult as possible.

A recent trend in microcontrollers is the introduction of 32-bit CPUs that are becoming cheaper and more easily available than 8-bit CPUs, in addition to being more easily programmable. The flash memory size is probably easier to grow than other parameters in microcontrollers. Flash memory size is not expected to be the most significant limiting factor. Before picking a platform, developers should also plan for firmware updates. This would essentially mean that the platform should at least have a flash memory size of the total code size * 2, plus some space for buffer.

The situation is less clear with regards to the amount of CPU power needed to run the algorithms. The demonstrated speeds are sufficient for many applications. For instance, a sensor that wakes up every now and then can likely spend a fraction of a second, or even spend multiple seconds in some cases, for the computation of a signature for the message that it is about to send. Most applications that use protocols such as DTLS that use public key cryptography only at the beginning of the session would also be fine with any of these execution times.

Yet, with reasonably long key sizes the execution times are in the seconds, dozens of seconds, or even longer. For some applications

this is too long. Nevertheless, these algorithms can successfully be employed in resource-constrained devices for the following reasons:

- With the right selection of algorithms and libraries, the execution times can actually be very small (less than 500 ms).
- As discussed in [wiman], in general the power requirements necessary to turn the radio on/off and sending or receiving messages are far bigger than those needed to execute cryptographic operations. While there are newer radios that significantly lower the energy consumption of sending and receiving messages, there is no good reason to choose platforms that do not provide sufficient computing power to run the necessary cryptographic operations.
- Commercial libraries and the use of full potential for various optimizations will provide a better result than what we arrived at in this memo.
- Using public-key cryptography only at the beginning of a session will reduce the per-packet processing times significantly.

While we did not do an exhaustive performance evaluation of asymmetric key pair generation on resource-constrained devices, we did note that it is possible for such devices to generate a new key pair. Given that this operation would only occur in rare circumstances (such as a factory reset or ownership change) and its potential privacy benefits, developers should provide mechanisms for generating new identities. It is however extremely important to note that the security of this operation relies on access to cryptographic-quality randomness.

8.2. Freshness

In our architecture, if implemented as described thus far, messages along with their signatures sent from the sensors to the publish-subscribe broker can be recorded and replayed by an eavesdropper. The publish-subscribe broker has no mechanism to distinguish previously received packets from those that are retransmitted by the sender or replayed by an eavesdropper. Therefore, it is essential for the smart objects to ensure that data updates include a freshness indicator. However, ensuring freshness on constrained devices can be non-trivial because of several reasons which include:

- Communication is mostly unidirectional to save energy.
- Internal clocks might not be accurate and may be reset several times during the operational phase of the smart object.

- Network time synchronization protocols such as Network Time Protocol (NTP) [RFC5905] are resource intensive and therefore may be undesirable in many smart object networks.

There are several different methods that can be used in our architecture for replay protection. The selection of the appropriate choice depends on the actual deployment scenario.

Including sequence numbers in signed messages can provide an effective method of replay protection. The publish-subscribe broker should verify the sequence number of each incoming message and accept it only if it is greater than the highest previously seen sequence number. The publish-subscribe broker drops any packet with a sequence number that has already been received or if the received sequence number is greater than the highest previously seen sequence number by an amount larger than the preset threshold.

Sequence numbers can wrap around at their maximum value and, therefore, it is essential to ensure that sequence numbers are sufficiently long. However, including long sequence numbers in packets can increase the network traffic originating from the sensor and can thus decrease its energy efficiency. To overcome the problem of long sequence numbers, we can use a scheme similar to that of Huang [huang], where the sender and receiver maintain and sign long sequence numbers of equal bit-lengths but they transmit only the least significant bits.

It is important for the smart object to write the sequence number into the permanent flash memory after each increment and before it is included in the message to be transmitted. This ensures that the sensor can obtain the last sequence number it had intended to send in case of a reset or a power failure. However, the sensor and the publish-subscribe broker can still end up in a discordant state where the sequence number received by the publish-subscribe broker exceeds the expected sequence number by an amount greater than the preset threshold. This may happen because of a prolonged network outage or if the publish-subscribe broker experiences a power failure for some reason. Therefore it is essential for sensors that normally send Non-Confirmable data updates to send some Confirmable updates and re-synchronize with the publish-subscribe broker if a reset message is received. The sensors re-synchronize by sending a new registration message with the current sequence number.

Although sequence numbers protect the system from replay attacks, a publish-subscribe broker has no mechanism to determine the time at which updates were created by the sensor. Moreover, if sequence numbers are the only freshness indicator used, a malicious eavesdropper can induce inordinate delays to the communication of

signed updates by buffering messages. It may be important in certain smart object networks for sensors to send data updates which include timestamps to allow the publish-subscribe broker to determine the time when the update was created. For example, when the publish-subscribe broker is collecting temperature data, it may be necessary to know when exactly the temperature measurement was made by the sensor. A simple solution to this problem is for the publish-subscribe broker to assume that the data object was created when it receives the update. In a relatively reliable network with low RTT, it can be acceptable to make such an assumption. However most networks are susceptible to packet loss and hostile attacks making this assumption unsustainable.

Depending on the hardware used by the smart objects, they may have access to accurate hardware clocks which can be used to include timestamps in the signed updates. These timestamps are included in addition to sequence numbers. The clock time in the smart objects can be set by the manufacturer or the current time can be communicated by the publish-subscribe broker during the registration phase. However, these approaches require the smart objects to either rely on the long-term accuracy of the clock set by the manufacturer or to trust the publish-subscribe broker thereby increasing the potential vulnerability of the system. The smart objects could also obtain the current time from NTP, but this may consume additional energy and give rise to security issues discussed in [RFC5905]. The smart objects could also have access to a mobile network or the Global Positioning System (GPS), and they can be used obtain the current time. Finally, if the sensors need to co-ordinate their sleep cycles, or if the publish-subscribe broker computes an average or mean of updates collected from multiple smart objects, it is important for the network nodes to synchronize the time among them. This can be done by using existing synchronization schemes.

8.3. Layering

It would be useful to select just one layer where security is provided at. Otherwise a simple device needs to implement multiple security mechanisms. While some code can probably be shared across such implementations (like algorithms), it is likely that most of the code involving the actual protocol machinery cannot. Looking at the different layers, here are the choices and their implications:

link layer

: This is probably the most common solution today. The biggest benefits of this choice of layer are that security services are commonly available (WLAN secrets, cellular SIM cards, etc.) and that their application protects the entire communications.

The main drawback is that there is no security beyond the first hop. This can be problematic, e.g., in many devices that communicate to a server in the Internet. A Withings scale [Withings], for instance, can support WLAN security but without some level of end-to-end security, it would be difficult to prevent fraudulent data submissions to the servers.

Another drawback is that some commonly implemented link layer security designs use group secrets. This allows any device within the local network (e.g., an infected laptop) to attack the communications.

network layer

There are a number of solutions in this space, and many new ones and variations thereof being proposed: IPsec, PANA, and so on. In general, these solutions have similar characteristics to those in the transport layer: they work across forwarding hops but only as far as to the next middlebox or application entity. There is plenty of existing solutions and designs.

Experience has shown that it is difficult to control IP layer entities from an application process. While this is theoretically easy, in practice the necessary APIs do not exist. For instance, most IPsec software has been built for the VPN use case, and is difficult or impossible to tweak to be used on a per-application basis. As a result, the authors are not particularly enthusiastic about recommending these solutions.

transport and application layer

This is another popular solution along with link layer designs. TLS with HTTP (HTTPS) and DTLS with CoAP are examples of solutions in this space, and have been proven to work well. These solutions are typically easy to take into use in an application, without assuming anything from the underlying OS, and they are easy to control as needed by the applications. The main drawback is that generally speaking, these solutions only run as far as the next application level entity. And even for this case, HTTPS can be made to work through proxies, so this limit is not unsolvable. Another drawback is that attacks on link layer, network layer and in some cases, transport layer, can not be protected against. However, if the upper layers have been protected, such attacks can at most result in a denial-of-service. Since denial-of-service can often be caused anyway, it is not clear if this is a real drawback.

data object layer

This solution does not protect any of the protocol layers, but protects individual data elements being sent. It works particularly well when there are multiple application layer entities on the path of the data. Smart object networks are likely to employ such entities for storage, filtering, aggregation and other reasons, and as such, an end-to-end solution is the only one that can protect the actual data.

The downside is that the lower layers are not protected. But again, as long as the data is protected and checked upon every time it passes through an application level entity, it is not clear that there are attacks beyond denial-of-service.

The main question mark is whether this type of a solution provides sufficient advantages over the more commonly implemented transport and application layer solutions.

8.4. Symmetric vs. Asymmetric Crypto

The second trade-off that is worth discussing is the use of plain asymmetric cryptographic mechanisms, plain symmetric cryptographic mechanisms, or some mixture thereof.

Contrary to popular cryptographic community beliefs, a symmetric cryptographic solution can be deployed in large scale. In fact, one of the largest deployment of cryptographic security, the cellular network authentication system, uses SIM cards that are based on symmetric secrets. In contrast, public key systems have yet to show ability to scale to hundreds of millions of devices, let alone billions. But the authors do not believe scaling is an important differentiator when comparing the solutions.

As can be seen from the Section 6, the time needed to calculate some of the asymmetric cryptographic operations with reasonable key lengths can be significant. There are two contrary observations that can be made from this. First, recent wisdom indicates that computing power on resource-constrained devices is far cheaper than transmission power [wiman], and keeps on becoming more efficient very quickly. From this we can conclude that the sufficient CPU is or at least will be easily available.

But the other observation is that when there are very costly asymmetric operations, doing a key exchange followed by the use of generated symmetric keys would make sense. This model works very well for DTLS and other transport layer solutions, but works less well for data object security, particularly when the number of communicating entities is not exactly two.

9. Summary

This document makes several security recommendations based on our implementation experience. We summarize some of the important ones here:

- Developers and product designers should choose the hardware after determining the security requirements for their application scenario.
- Elliptic Curve Cryptography (ECC) outperforms RSA based operations and therefore it is recommended for resource-constrained devices.
- Cryptographic-quality randomness is needed for many security protocols. Developers and vendors should ensure that the sufficient randomness is available for security critical tasks.
- 32-bit microcontrollers are much more easily available, at lower costs and are more power efficient. Therefore, real-world deployments are better off using 32-bit microcontrollers.
- Developers should provide mechanisms for devices to generate new identities at appropriate times during their lifecycle. For example, after a factory reset or an ownership handover.
- Planning for firmware updates is important. The hardware platform chosen should at least have a flash memory size of the total code size * 2, plus some space for buffer.

10. Security Considerations

This entire memo deals with security issues.

11. IANA Considerations

There are no IANA impacts in this memo.

12. Informative references

[arduino-uno]
Arduino, "Arduino Uno", September 2015, <http://arduino.cc/en/Main/arduinoBoardUno>.

[armecdsa]
Tschofenig, H. and M. Pegourie-Gonnard, "Performance Investigations", March 2015, <https://www.ietf.org/proceedings/92/slides/slides-92-lwig-3.pdf>.

[avr-crypto-lib]
AVR-CRYPTO-LIB, "AVR-CRYPTO-LIB", September 2015, <http://www.das-labor.org/wiki/AVR-Crypto-Lib/en>.

[avr-cryptolib]
Van der Laan, E., "AVR CRYPTOLIB", September 2015, <http://www.emsign.nl/>.

[avrora] Titzer, Ben., "Avrora", September 2015, <http://compilers.cs.ucla.edu/avrora/>.

[freescale]
NXP, "Freescale FRDM-KL25Z", June 2017, <https://developer.mbed.org/platforms/KL25Z/>.

[hahmos] Hahm, O., Baccelli, E., Petersen, H., and N. Tsiftes, "Operating systems for low-end devices in the internet of things: a survey", IEEE Internet of Things Journal , 2016.

[huang] Huang, C., "Low-overhead freshness transmission in sensor networks", 2008.

[I-D.arkko-core-security-arch]
Arkko, J. and A. Keranen, "CoAP Security Architecture", draft-arkko-core-security-arch-00 (work in progress), July 2011.

[I-D.arkko-core-sleepy-sensors]
Arkko, J., Rissanen, H., Loreto, S., Turanyi, Z., and O. Novo, "Implementing Tiny COAP Sensors", draft-arkko-core-sleepy-sensors-01 (work in progress), July 2011.

[I-D.daniel-6lowpan-security-analysis]
Park, S., Kim, K., Haddad, W., Chakrabarti, S., and J. Laganier, "IPv6 over Low Power WPAN Security Analysis", draft-daniel-6lowpan-security-analysis-05 (work in progress), March 2011.

[I-D.ietf-core-coap-pubsub]
Koster, M., Keranen, A., and J. Jimenez, "Publish-Subscribe Broker for the Constrained Application Protocol (CoAP)", draft-ietf-core-coap-pubsub-03 (work in progress), February 2018.

[I-D.ietf-core-resource-directory]
Shelby, Z., Koster, M., Bormann, C., Stok, P., and C. Amsuess, "CoRE Resource Directory", draft-ietf-core-resource-directory-12 (work in progress), October 2017.

[I-D.ietf-core-senml]
Jennings, C., Shelby, Z., Arkko, J., Keranen, A., and C. Bormann, "Media Types for Sensor Measurement Lists (SenML)", draft-ietf-core-senml-12 (work in progress), December 2017.

[I-D.irtf-t2trg-iot-seccons]
Garcia-Morchon, O., Kumar, S., and M. Sethi, "State-of-the-Art and Challenges for the Internet of Things Security", draft-irtf-t2trg-iot-seccons-11 (work in progress), February 2018.

[I-D.moskowitz-hip-dex]
Moskowitz, R. and R. Hummen, "HIP Diet EXchange (DEX)", draft-moskowitz-hip-dex-05 (work in progress), January 2016.

[I-D.sarikaya-t2trg-sbootstrapping]
Sarikaya, B., Sethi, M., and A. Sangi, "Secure IoT Bootstrapping: A Survey", draft-sarikaya-t2trg-sbootstrapping-03 (work in progress), February 2017.

[matrix-ssl]
PeerSec Networks, "Matrix SSL", September 2015, <http://www.matrixssl.org/>.

[mbed] ARM, "mbed TLS", May 2017, <https://www.mbed.com/en/technologies/security/mbed-tls/>.

[micronacl]
MicroNaCl, "The Networking and Cryptography library for microcontrollers", <http://munacl.cryptojedi.org/>.

[mosdorf] Mosdorf, M. and W. Zabolotny, "Implementation of elliptic curve cryptography for 8 bit and 32 bit embedded systems time efficiency and power consumption analysis", Pomiary Automatyka Kontrola , 2010.

[nacl] NaCl, "Networking and Cryptography library", <http://nacl.cr.yp.to/>.

[naclavr] Hutter, M. and P. Schwabe, "NaCl on 8-Bit AVR Microcontrollers", International Conference on Cryptology in Africa , Springer Berlin Heidelberg , 2013.

[nesC] Gay, D., Levis, P., von Behren, R., Welsh, M., Brewer, E., and D. Culler, "The nesC language: A holistic approach to networked embedded systems", ACM SIGPLAN Notices , 2014.

[nordic] Nordic Semiconductor, "nRF52832 Product Specification", June 2017, <http://infocenter.nordicsemi.com/pdf/nRF52832_PS_v1.3.pdf>.

[relic-toolkit] Aranha, D. and C. Gouv, "Relic Toolkit", September 2015, <http://code.google.com/p/relic-toolkit/>.

[RFC3748] Aboba, B., Blunk, L., Vollbrecht, J., Carlson, J., and H. Levkowetz, Ed., "Extensible Authentication Protocol (EAP)", RFC 3748, DOI 10.17487/RFC3748, June 2004, <https://www.rfc-editor.org/info/rfc3748>.

[RFC3972] Aura, T., "Cryptographically Generated Addresses (CGA)", RFC 3972, DOI 10.17487/RFC3972, March 2005, <https://www.rfc-editor.org/info/rfc3972>.

[RFC4086] Eastlake 3rd, D., Schiller, J., and S. Crocker, "Randomness Requirements for Security", BCP 106, RFC 4086, DOI 10.17487/RFC4086, June 2005, <https://www.rfc-editor.org/info/rfc4086>.

[RFC4303] Kent, S., "IP Encapsulating Security Payload (ESP)", RFC 4303, DOI 10.17487/RFC4303, December 2005, <https://www.rfc-editor.org/info/rfc4303>.

[RFC5191] Forsberg, D., Ohba, Y., Ed., Patil, B., Tschofenig, H., and A. Yegin, "Protocol for Carrying Authentication for Network Access (PANA)", RFC 5191, DOI 10.17487/RFC5191, May 2008, <https://www.rfc-editor.org/info/rfc5191>.

[RFC5246] Dierks, T. and E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.2", RFC 5246, DOI 10.17487/RFC5246, August 2008, <https://www.rfc-editor.org/info/rfc5246>.

[RFC5406] Bellovin, S., "Guidelines for Specifying the Use of IPsec Version 2", BCP 146, RFC 5406, DOI 10.17487/RFC5406, February 2009, <https://www.rfc-editor.org/info/rfc5406>.

[RFC5905] Mills, D., Martin, J., Ed., Burbank, J., and W. Kasch, "Network Time Protocol Version 4: Protocol and Algorithms Specification", RFC 5905, DOI 10.17487/RFC5905, June 2010, <https://www.rfc-editor.org/info/rfc5905>.

[RFC6078] Camarillo, G. and J. Melen, "Host Identity Protocol (HIP) Immediate Carriage and Conveyance of Upper-Layer Protocol Signaling (HICCUPS)", RFC 6078, DOI 10.17487/RFC6078, January 2011, <https://www.rfc-editor.org/info/rfc6078>.

[RFC6347] Rescorla, E. and N. Modadugu, "Datagram Transport Layer Security Version 1.2", RFC 6347, DOI 10.17487/RFC6347, January 2012, <https://www.rfc-editor.org/info/rfc6347>.

[RFC6574] Tschofenig, H. and J. Arkko, "Report from the Smart Object Workshop", RFC 6574, DOI 10.17487/RFC6574, April 2012, <https://www.rfc-editor.org/info/rfc6574>.

[RFC6690] Shelby, Z., "Constrained RESTful Environments (CoRE) Link Format", RFC 6690, DOI 10.17487/RFC6690, August 2012, <https://www.rfc-editor.org/info/rfc6690>.

[RFC7230] Fielding, R., Ed. and J. Reschke, Ed., "Hypertext Transfer Protocol (HTTP/1.1): Message Syntax and Routing", RFC 7230, DOI 10.17487/RFC7230, June 2014, <https://www.rfc-editor.org/info/rfc7230>.

[RFC7252] Shelby, Z., Hartke, K., and C. Bormann, "The Constrained Application Protocol (CoAP)", RFC 7252, DOI 10.17487/RFC7252, June 2014, <https://www.rfc-editor.org/info/rfc7252>.

[RFC7296] Kaufman, C., Hoffman, P., Nir, Y., Eronen, P., and T. Kivinen, "Internet Key Exchange Protocol Version 2 (IKEv2)", STD 79, RFC 7296, DOI 10.17487/RFC7296, October 2014, <https://www.rfc-editor.org/info/rfc7296>.

[RFC7401] Moskowitz, R., Ed., Heer, T., Jokela, P., and T. Henderson, "Host Identity Protocol Version 2 (HIPv2)", RFC 7401, DOI 10.17487/RFC7401, April 2015, <https://www.rfc-editor.org/info/rfc7401>.

[RFC7515] Jones, M., Bradley, J., and N. Sakimura, "JSON Web Signature (JWS)", RFC 7515, DOI 10.17487/RFC7515, May 2015, <https://www.rfc-editor.org/info/rfc7515>.

[RFC7748] Langley, A., Hamburg, M., and S. Turner, "Elliptic Curves for Security", RFC 7748, DOI 10.17487/RFC7748, January 2016, <https://www.rfc-editor.org/info/rfc7748>.

[RFC7815] Kivinen, T., "Minimal Internet Key Exchange Version 2 (IKEv2) Initiator Implementation", RFC 7815, DOI 10.17487/RFC7815, March 2016, <https://www.rfc-editor.org/info/rfc7815>.

[RFC8032] Josefsson, S. and I. Liusvaara, "Edwards-Curve Digital Signature Algorithm (EdDSA)", RFC 8032, DOI 10.17487/RFC8032, January 2017, <https://www.rfc-editor.org/info/rfc8032>.

[RFC8152] Schaad, J., "CBOR Object Signing and Encryption (COSE)", RFC 8152, DOI 10.17487/RFC8152, July 2017, <https://www.rfc-editor.org/info/rfc8152>.

[rsa-8bit] Gura, N., Patel, A., Wander, A., Eberle, H., and S. Shantz, "Comparing Elliptic Curve Cryptography and RSA on 8-bit CPUs", 2010.

[rsa-high-speed] Koc, C., "High-Speed RSA Implementation", November 1994, <http://cs.ucsb.edu/~koc/docs/r01.pdf>.

[sec2ecc] Certicom Research, "SEC 2: Recommended Elliptic Curve Domain Parameters", 2000.

[stnucleo] STMicroelectronics, "NUCLEO-F091RC", June 2017, <http://www.st.com/en/evaluation-tools/nucleo-f091rc.html/>.

[tinyecc] North Carolina State University and North Carolina State University, "TinyECC", 2008, <http://discovery.csc.ncsu.edu/software/TinyECC/>.

[wiman] Margi, C., Oliveira, B., Sousa, G., Simplicio, M., Paulo, S., Carvalho, T., Naslund, M., and R. Gold, "Impact of Operating Systems on Wireless Sensor Networks (Security) Applications and Testbeds", International Conference on Computer Communication Networks (ICCCN'2010) / IEEE International Workshop on Wireless Mesh and Ad Hoc Networks (WiMAN 2010) , 2010.

[wiselib] Baumgartner, T., Chatzigiannakis, I., Fekete, S., Koninis, C., Kroller, A., and A. Pyrgelis, "Wiselib", 2010, <www.wiselib.org/>.

[Withings]
Withings, "The Withings scale", February 2012,
<http://www.withings.com/en/bodyscale>.

Appendix A. Acknowledgments

The authors would like to thank Mats Naslund, Salvatore Loreto, Bob Moskowitz, Oscar Novo, Vlasios Tsiatsis, Daoyuan Li, Muhammad Waqas, Eric Rescorla and Tero Kivinen for interesting discussions in this problem space. The authors would also like to thank Diego Aranha for helping with the relic-toolkit configurations and Tobias Baumgartner for helping with questions regarding wiselib.

Tim Chown, Samita Chakrabarti, Christian Huitema, Dan Romascanu, Eric Vyncke, and Emmanuel Baccelli provided valuable comments that helped us improve the final version of this document.

Authors' Addresses

Mohit Sethi
Ericsson
Jorvas 02420
Finland

EMail: mohit@piuha.net

Jari Arkko
Ericsson
Jorvas 02420
Finland

EMail: jari.arkko@piuha.net

Ari Keranen
Ericsson
Jorvas 02420
Finland

EMail: ari.keranen@ericsson.com

Heidi-Maria Back
Nokia
Helsinki 00181
Finland

EMail: heidi.back@nokia.com