Privacy-Preserving Pseudonyms for LoRaWAN

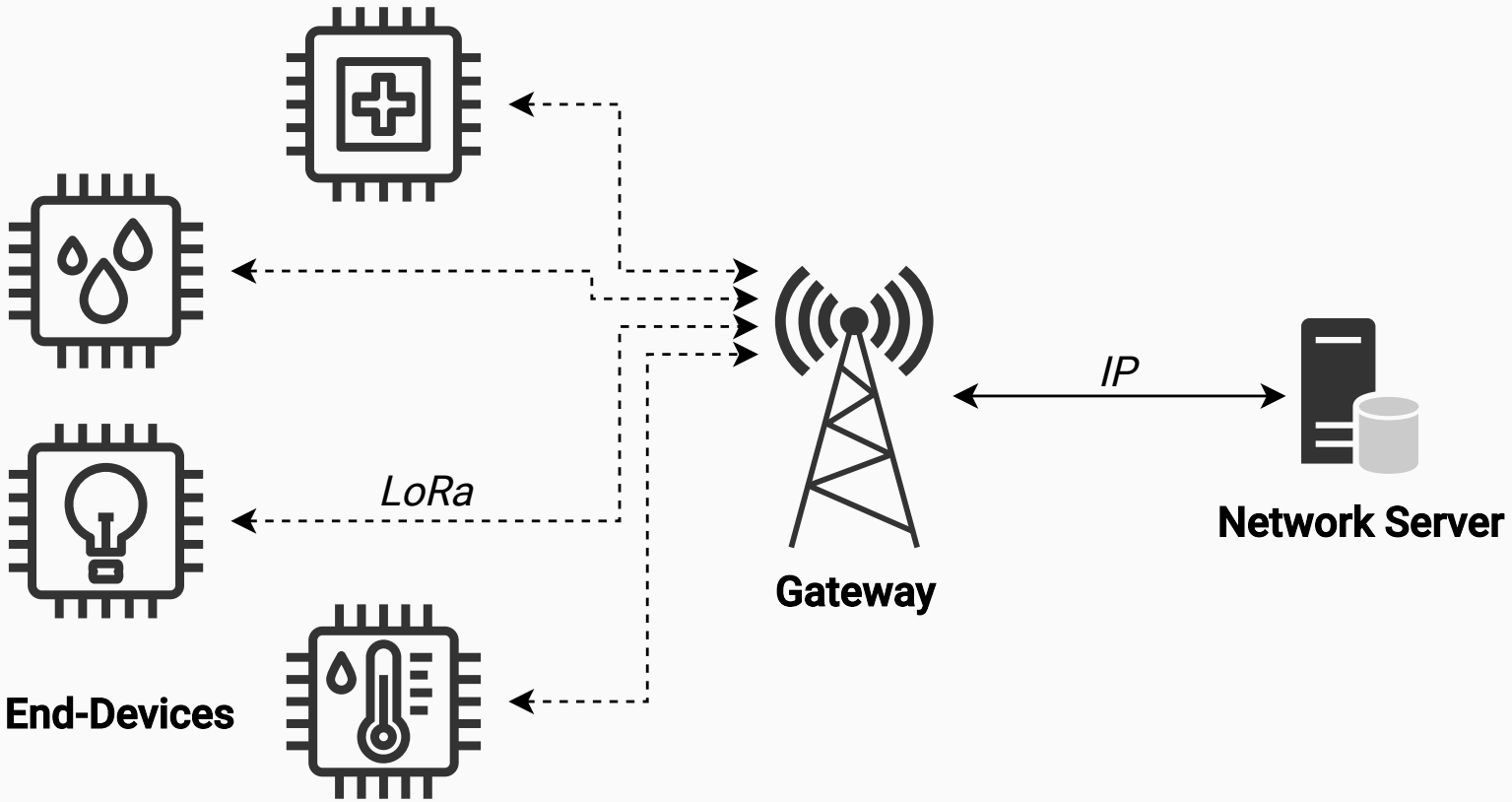
ACM WiSec'24

Samuel Pélissier*, Jan Aalmoes*
Abhishek Kumar Mishra*, Mathieu Cunche*
Vincent Roca**, Didier Donsez***

*: INSA-Lyon, Inria, CITI Lab.
**: University Grenoble Alpes, Inria
***: University Grenoble Alpes, LIG

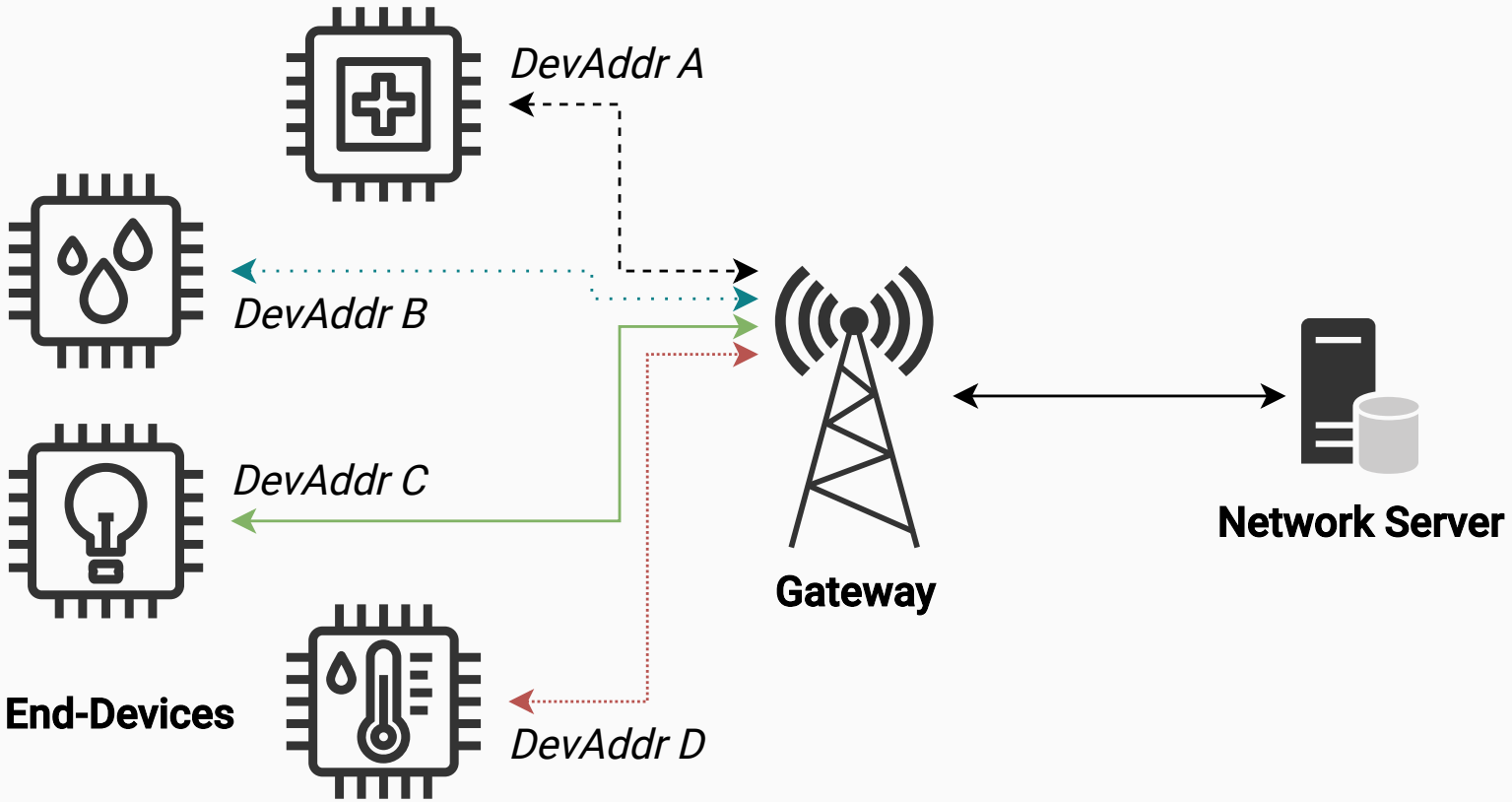
LoRaWAN

- **Low Power Wide Area Network** (LPWAN)
- **Lo**ng **Ra**nge modulation
- Long battery life and low energy consumption
- Mainly `uplink` communication (device to server)

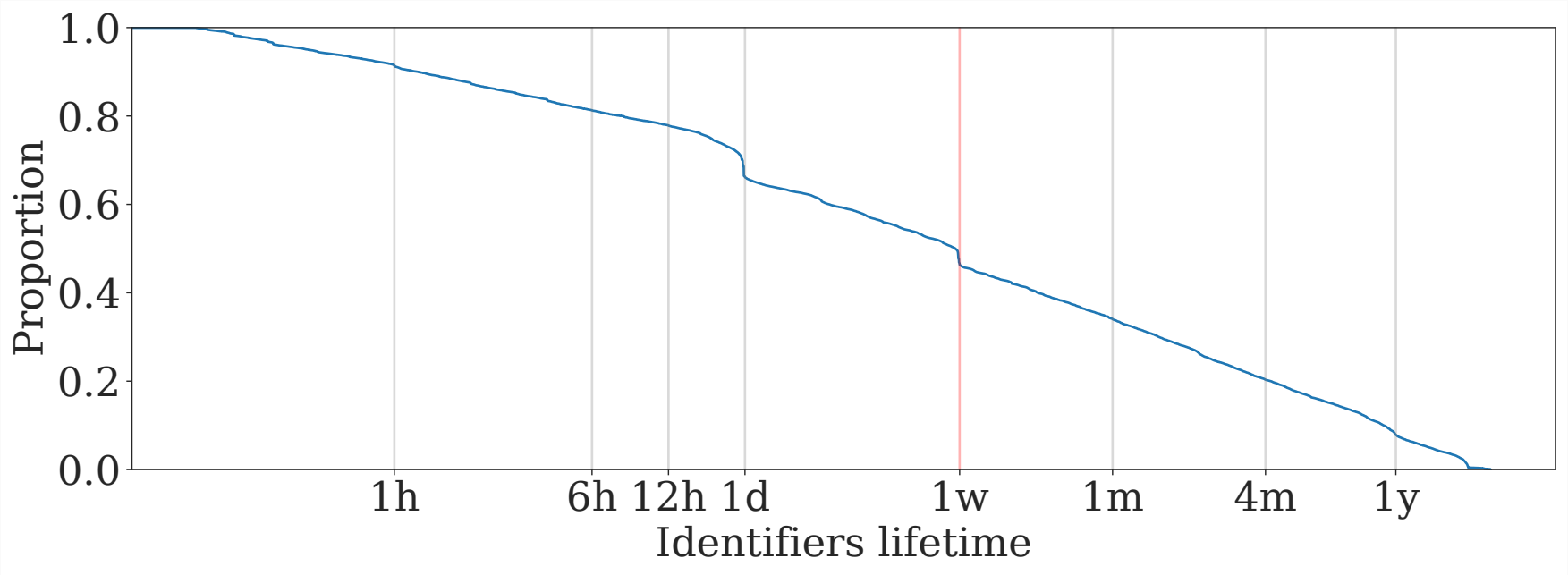


LoRaWAN

- **Low Power Wide Area Network** (LPWAN)
- **Lo**ng **Ra**nge modulation
- Long battery life and low energy consumption
- Mainly `uplink` communication (device to server)



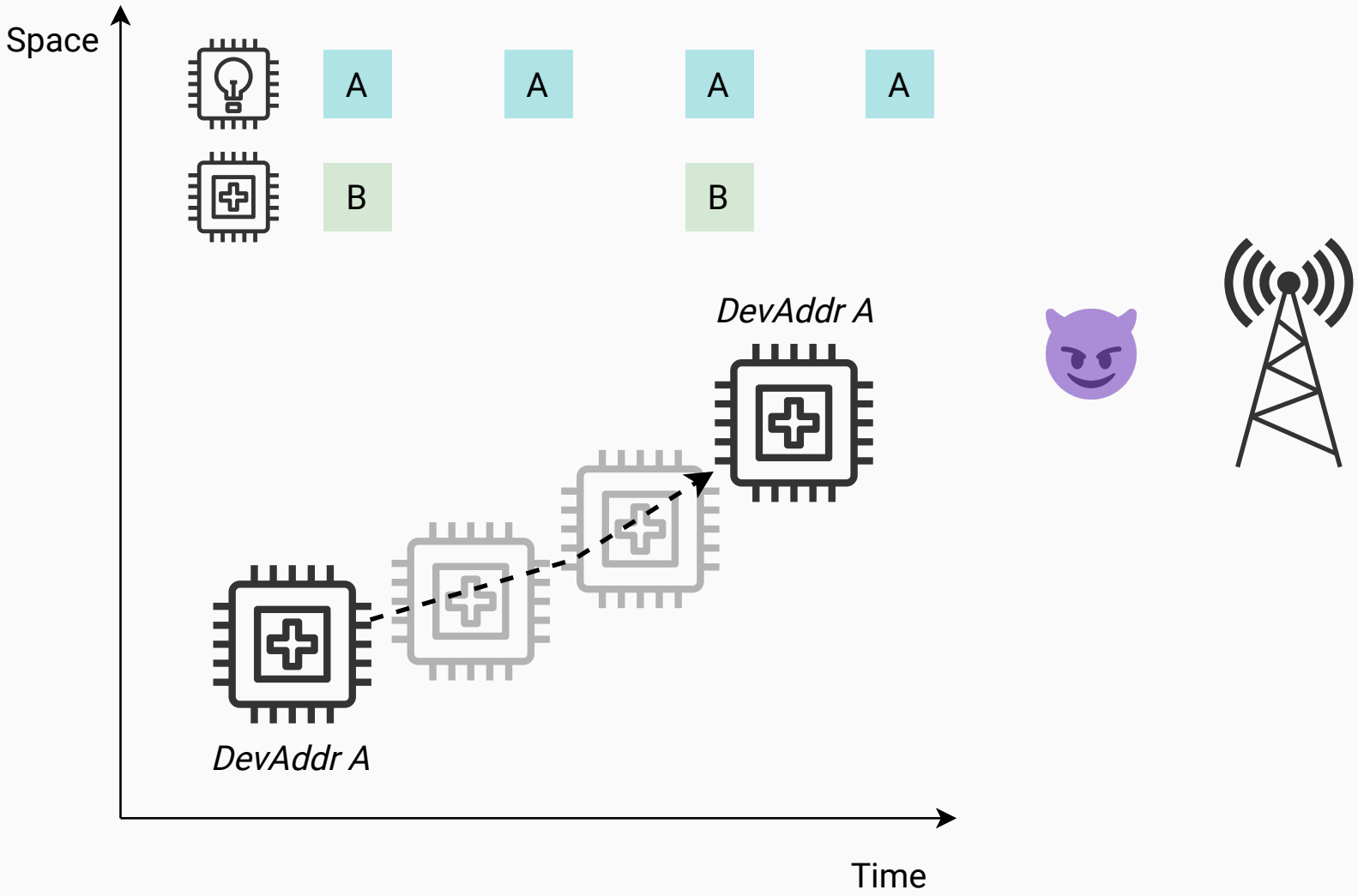
Devices leverage a stable DevAddr for long periods of time.



47% of devices can be tracked for more than one week via their `DevAddr`.

DevAddr stability

Stable DevAddr can be exploited for **activity inference** or **tracking**.



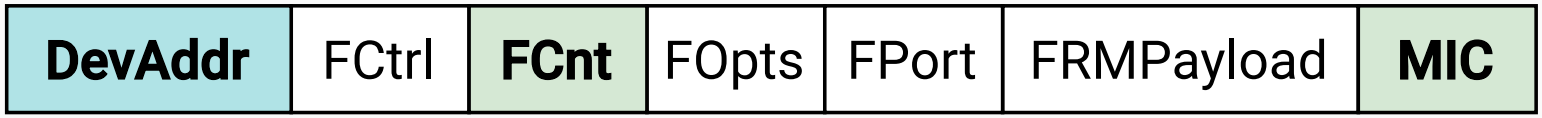
Needing LoRaWAN-tuned pseudonyms

A solution to protect privacy: temporary pseudonyms.

Existing approaches (in Bluetooth or Wi-Fi) need to be revisited for LoRaWAN:

- **Mainly** `uplink` **messages** (device to server), with occasional `downlink`;
- **High packet loss** (reported up to 40%);
- **Low energy consumption** (10 years on battery, optimized payload);
- **Sparse memory** (as low as 8*kB*);
- **Limited address space: 7 to 25 bits**, vs 48 bits for BLE and Wi-Fi.

LoRaWAN frame format



`DevAddr`

- Session identifier
- Set by the Network Server (join process)

`FCnt`

- Frame counter (starts at 0, incremented each message)
- Should be hidden to avoid tracking

`MIC`

- Message Integrity Code
- Computed using AES-128 CMAC and keys shared with the Network Server
- Can be used to identify devices

Properties

Privacy-preserving pseudonyms in LoRaWAN should satisfy:

$\mathcal{P}_1$: **Unlinkability**

$\mathcal{P}_2$: **Minimal communication overhead**

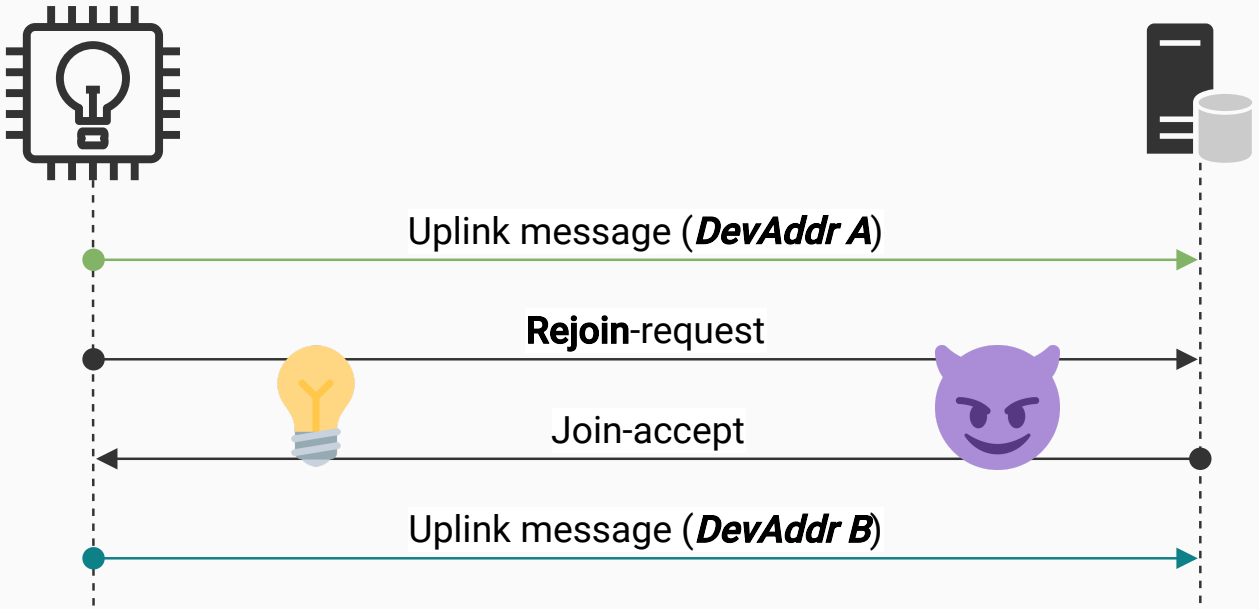
$\mathcal{P}_3$: **Legacy support**

$\mathcal{P}_4$: **Low computation/memory overhead**

$\mathcal{P}_5$: **Reliability**

Legacy schemes: leveraging the rejoin process

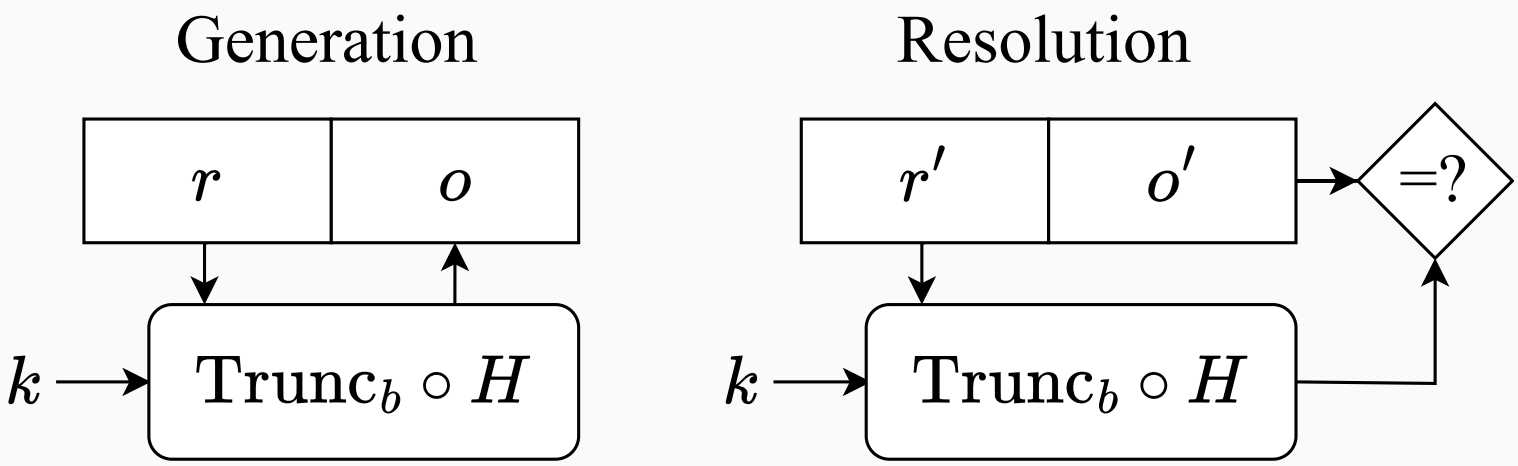
There is no satisfactory solution in the current LoRaWAN standard.



- Communication requires a lot of energy (10^9 times more than encryption)
- It generates exploitable metadata

Scheme 1: Resolvable pseudonyms in LoRaWAN

Generate pseudonyms by hashing a random value.

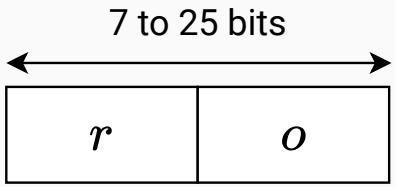


Built upon BLE's Resolvable random Private Addresses.

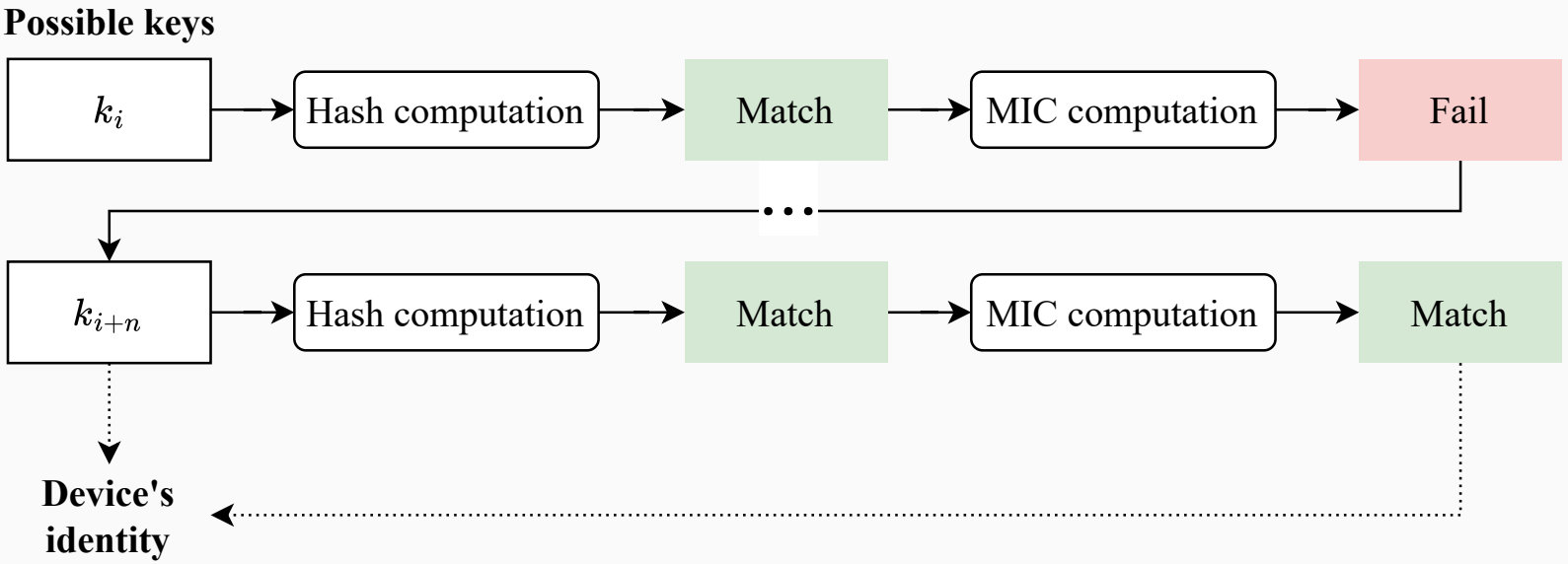
- *Shared resolution key*: Derived during join process
- *Hash function*: AES-128 CMAC
- *Need to hide metadata*: Encrypt `FCnt`

Scheme 1: Resolvable pseudonyms in LoRaWAN

Due to size constraints, **multiple keys** generate the **same truncated hash**.

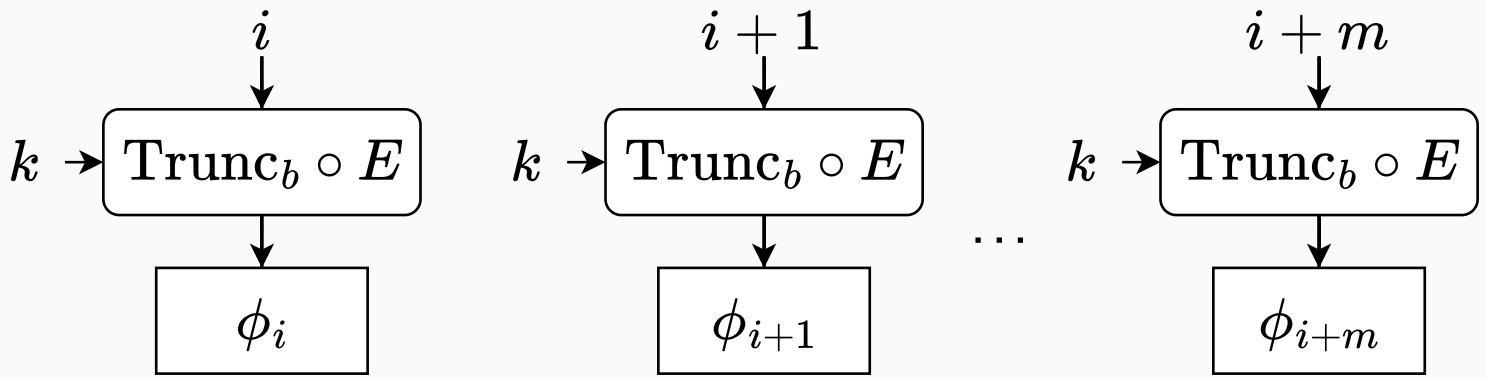


In case of collisions, **compute the MIC** to re-identify the sender.



Scheme 2: Sequential pseudonyms in LoRaWAN

Generate pseudonyms by encrypting a counter.

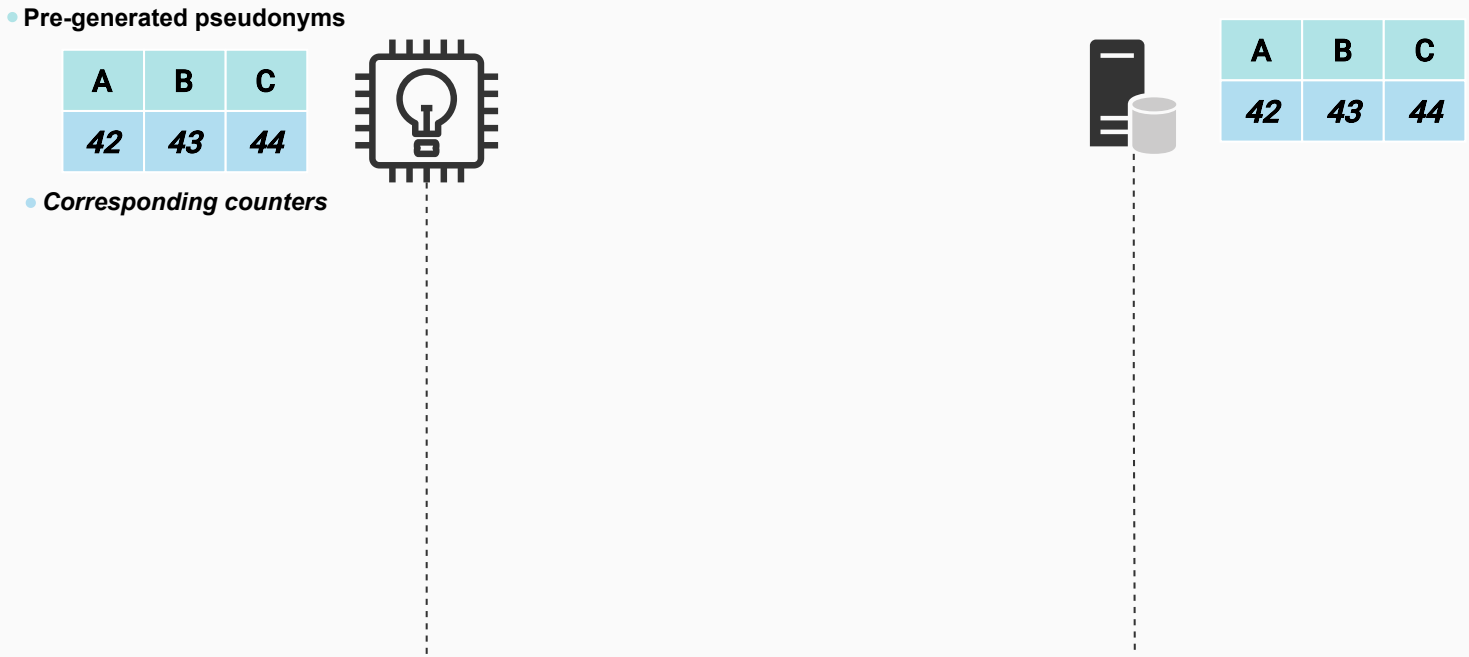


Built upon Shroud, proposed for Wi-Fi by Greenstein et al.

- *Shared resolution key*: Derived during join process
- *Encryption function*: AES-128 CTR
- *Need to hide metadata*: Re-use `FCnt`'s bits for the pseudonym

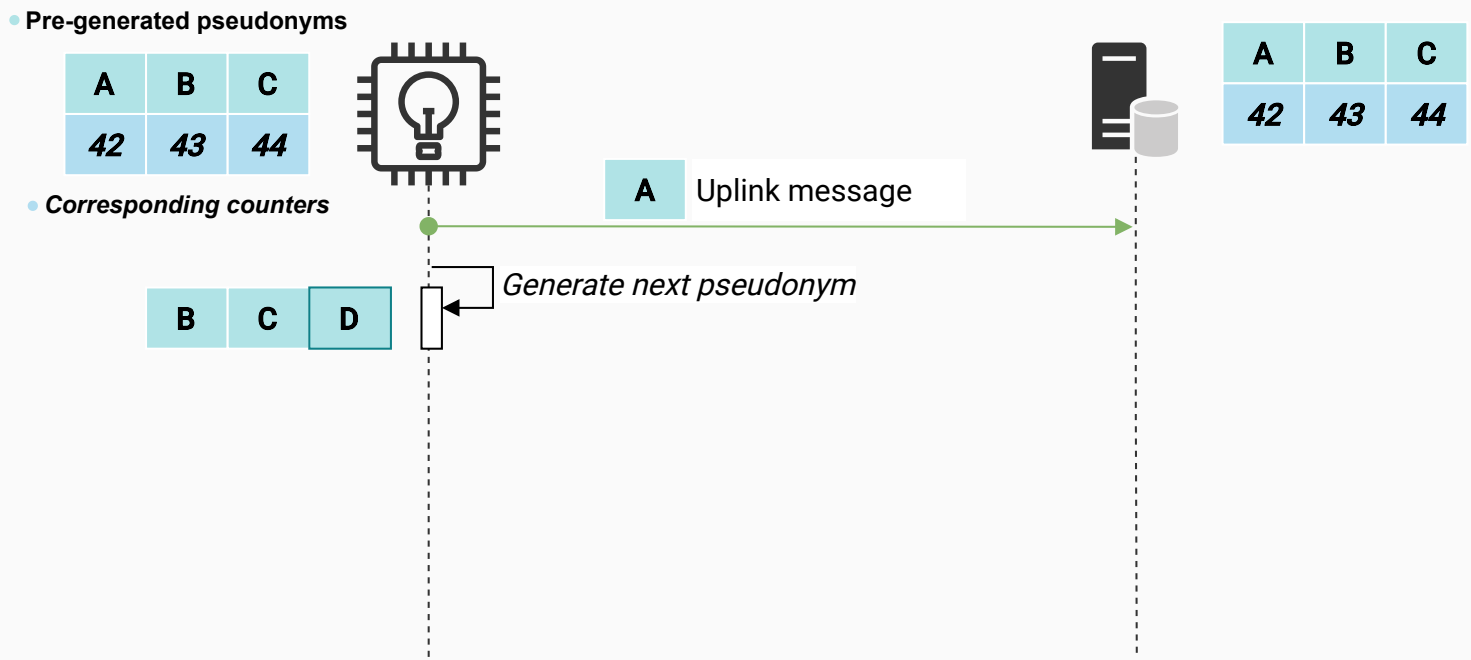
Total: 23 to 41 bits

Scheme 2: Sequential pseudonyms in LoRaWAN



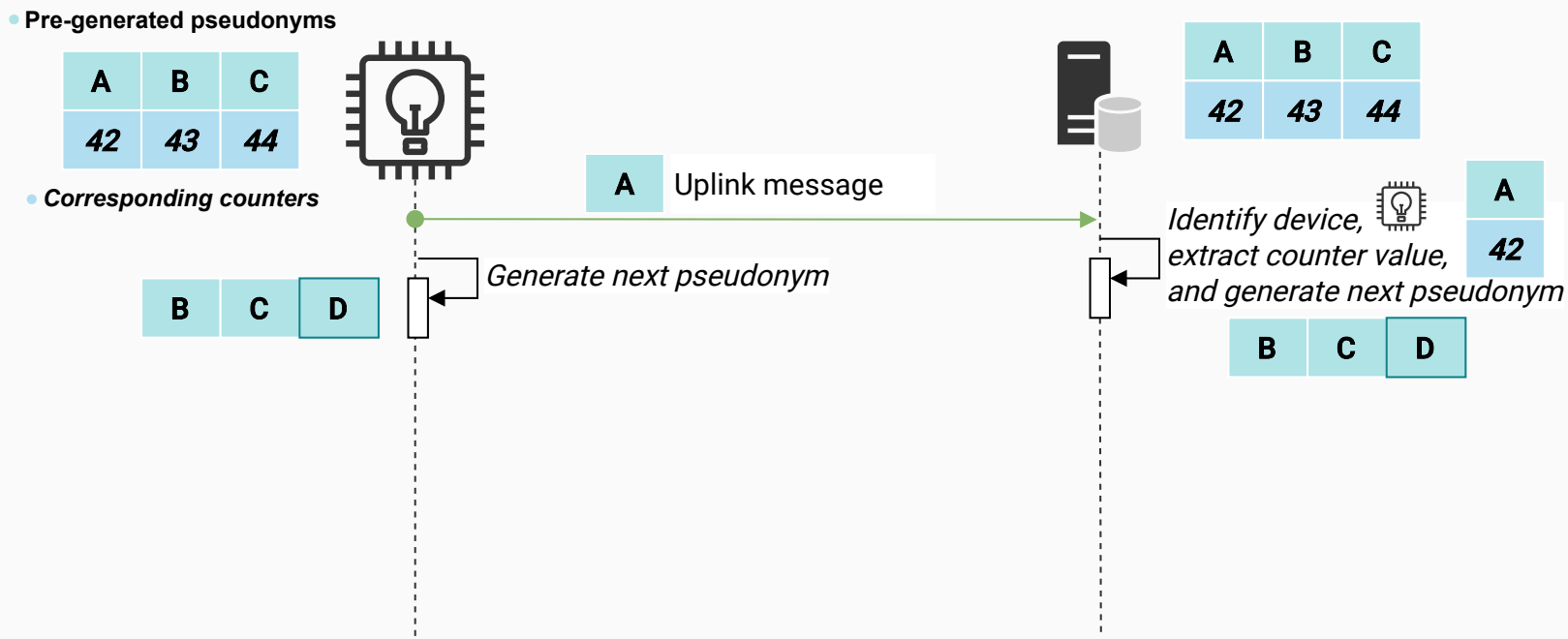
One device communicating with the Network Server, for 3 pre-generated pseudonyms.

Scheme 2: Sequential pseudonyms in LoRaWAN



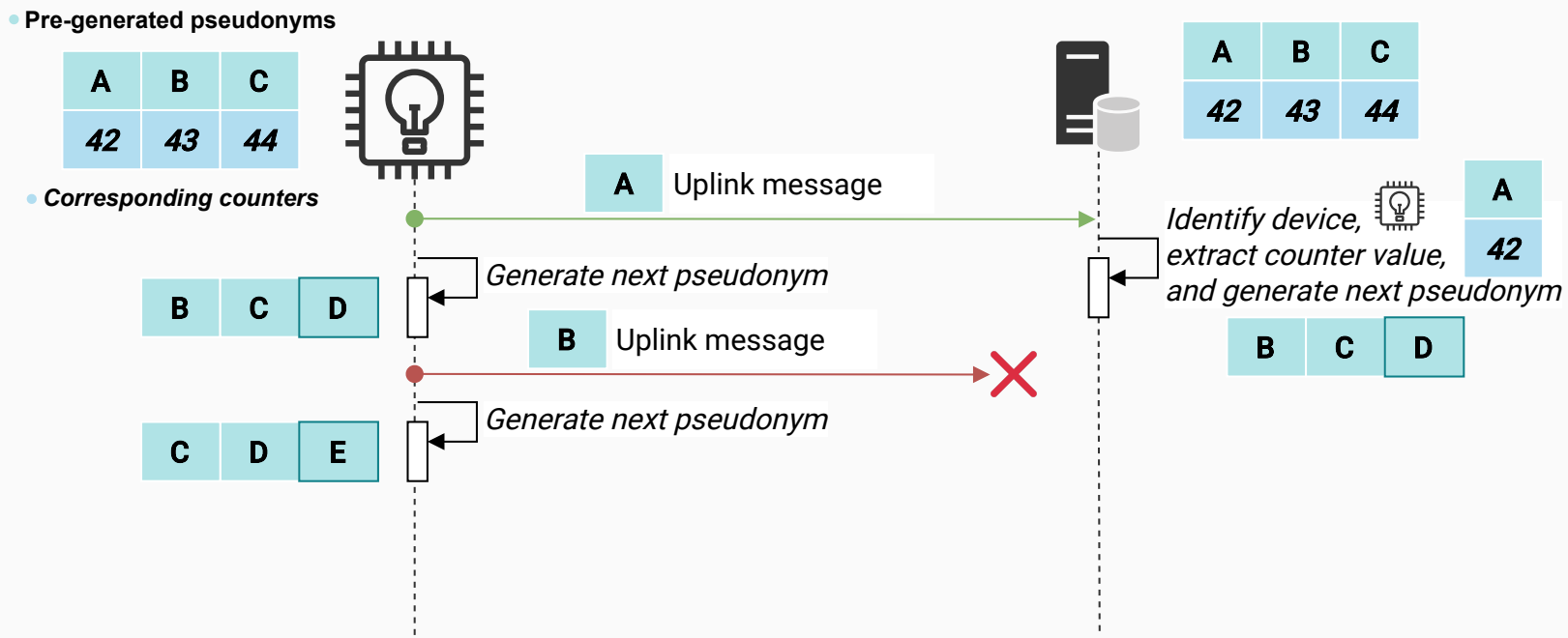
One device communicating with the Network Server, for 3 pre-generated pseudonyms.

Scheme 2: Sequential pseudonyms in LoRaWAN



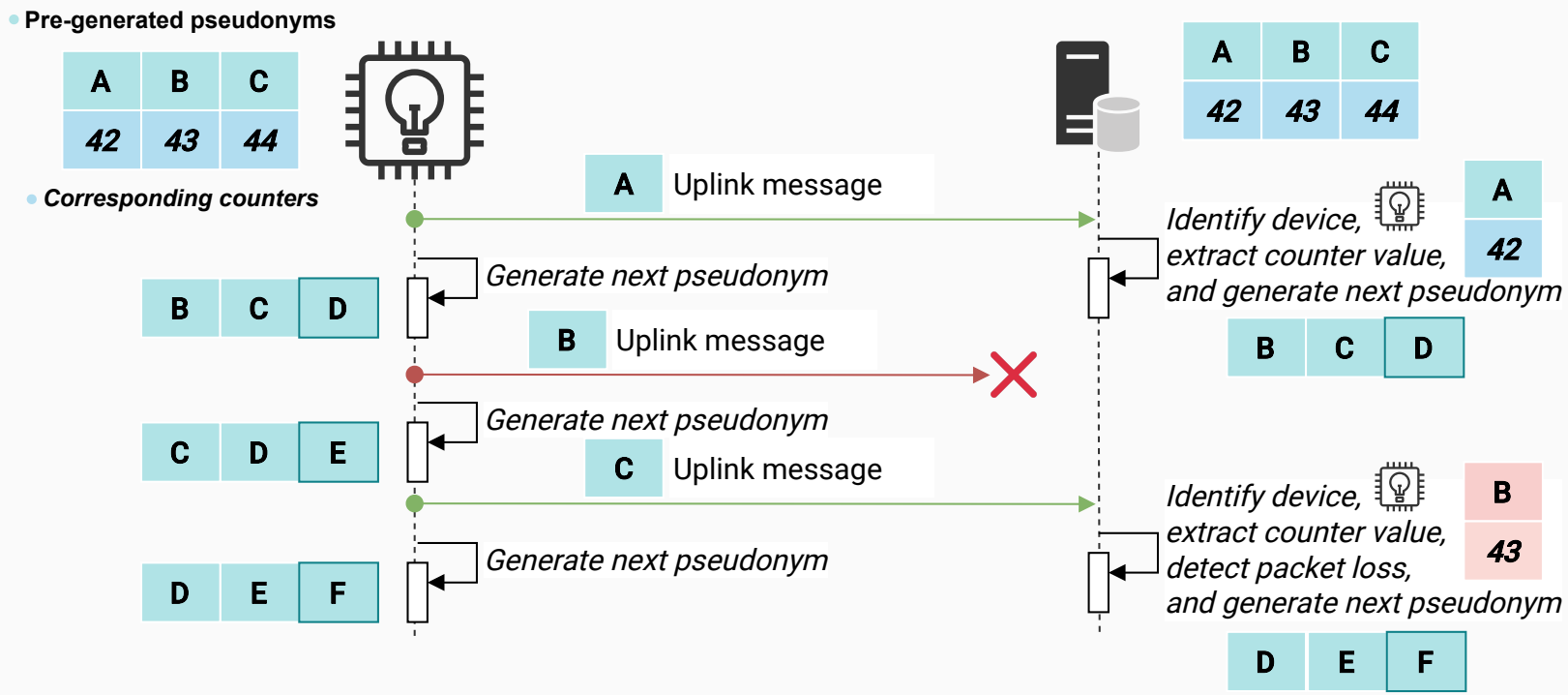
One device communicating with the Network Server, for 3 pre-generated pseudonyms.

Scheme 2: Sequential pseudonyms in LoRaWAN



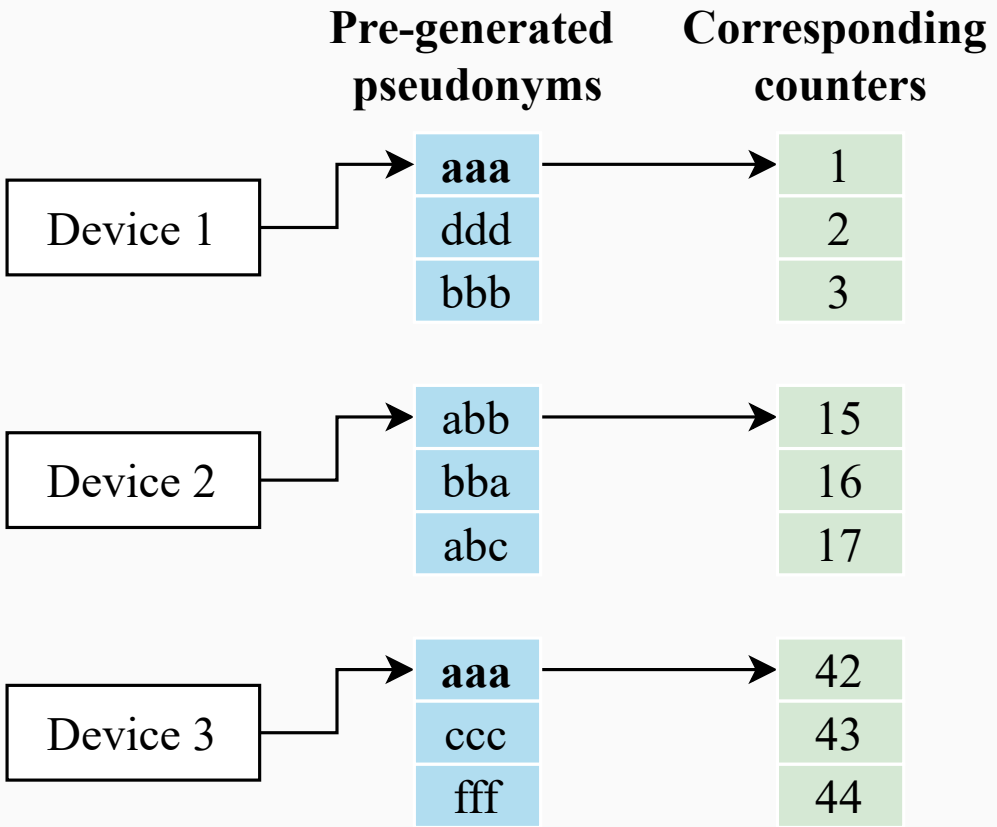
One device communicating with the Network Server, for 3 pre-generated pseudonyms.

Scheme 2: Sequential pseudonyms in LoRaWAN



One device communicating with the Network Server, for 3 pre-generated pseudonyms.

Scheme 2: Sequential pseudonyms in LoRaWAN



Multiple devices communicating with the Network Server, with 3 pre-generated pseudonyms each.

Evaluation: initial properties

Properties	Resolvable	Sequential
$\mathcal{P}_1$: Unlinkability	?	?
$\mathcal{P}_2$: Minimal communication overhead	?	?
$\mathcal{P}_3$: Legacy support	?	?
$\mathcal{P}_4$: Low computation/memory overhead	?	?
$\mathcal{P}_5$: Reliability	?	?

Unlinkability: guarantees of AES

Communication: no extra byte on the air

Legacy: major version field and header length maintained

Evaluation: initial properties

Properties	Resolvable	Sequential
$\mathcal{P}_1$: Unlinkability	✅	✅
$\mathcal{P}_2$: Minimal communication overhead	✅	✅
$\mathcal{P}_3$: Legacy support	✅	✅
$\mathcal{P}_4$: Low computation/memory overhead	?	?
$\mathcal{P}_5$: Reliability	?	?

- Unlinkability: guarantees of AES
- Communication: no extra byte on the air
- Legacy: major version field and header length maintained

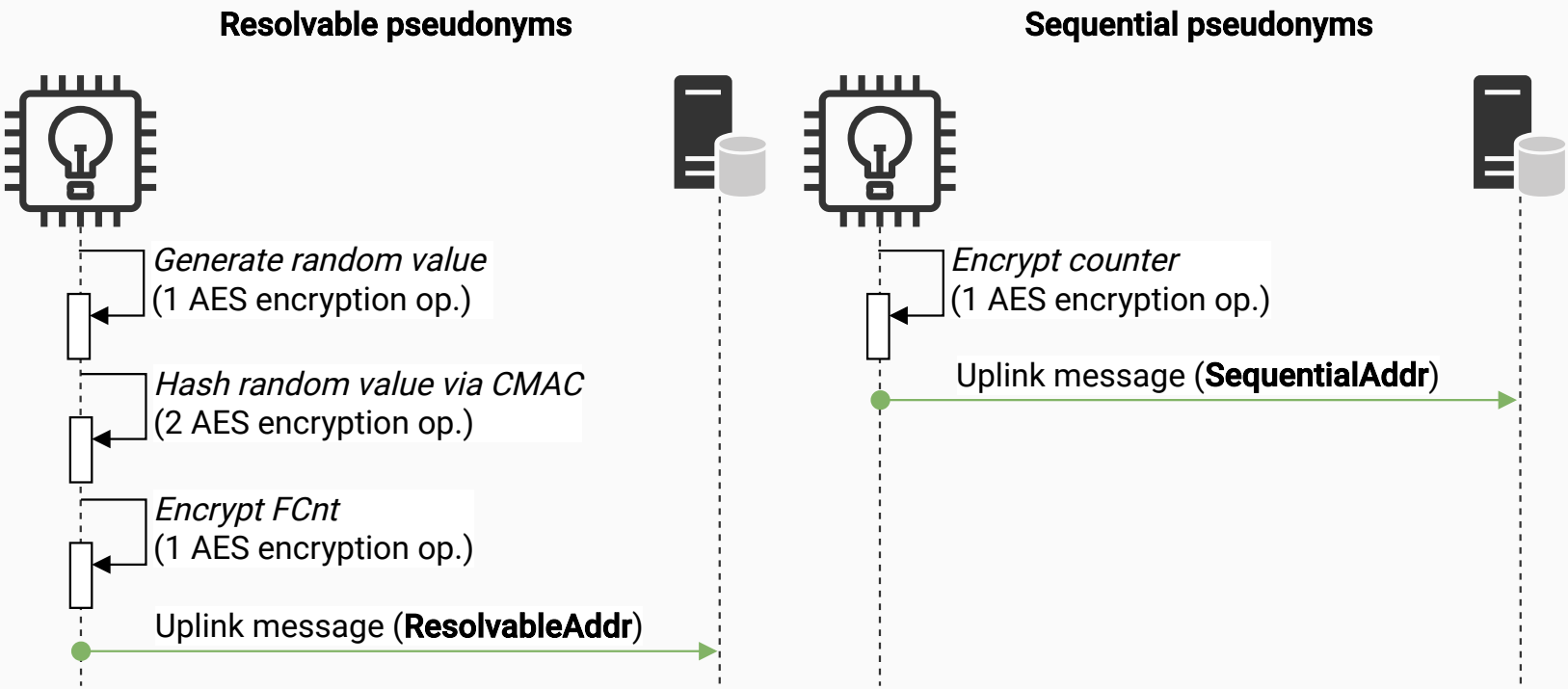
Evaluation: initial properties

Properties	Resolvable	Sequential
$\mathcal{P}_1$: Unlinkability	✅	✅
$\mathcal{P}_2$: Minimal communication overhead	✅	✅
$\mathcal{P}_3$: Legacy support	✅	✅
$\mathcal{P}_4$: Low computation/memory overhead	?	?
$\mathcal{P}_5$: Reliability	?	?

- Unlinkability: guarantees of AES
- Communication: no extra byte on the air
- Legacy: major version field and header length maintained

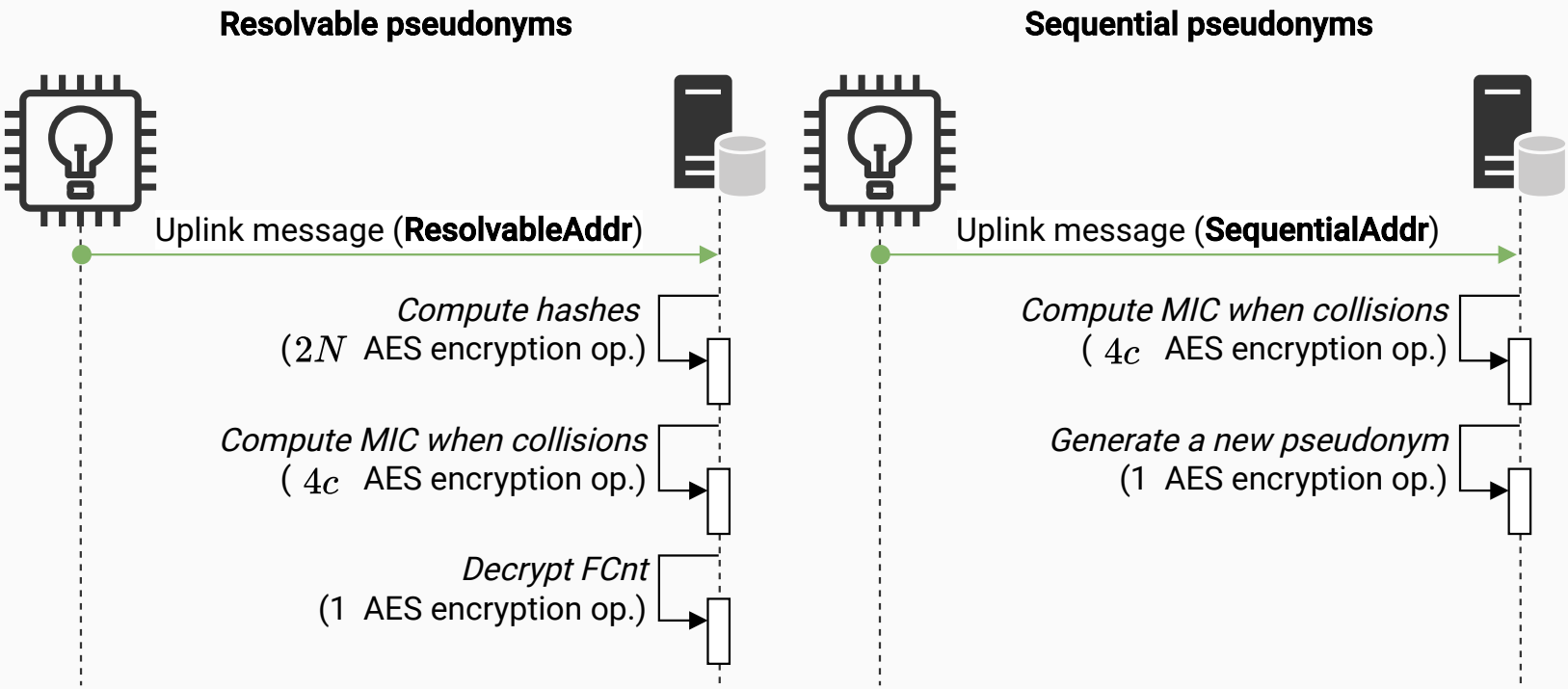
Evaluation (2): computation overhead

Device-side generation: resolvable pseudonyms have a higher overhead.



Evaluation (2): computation overhead

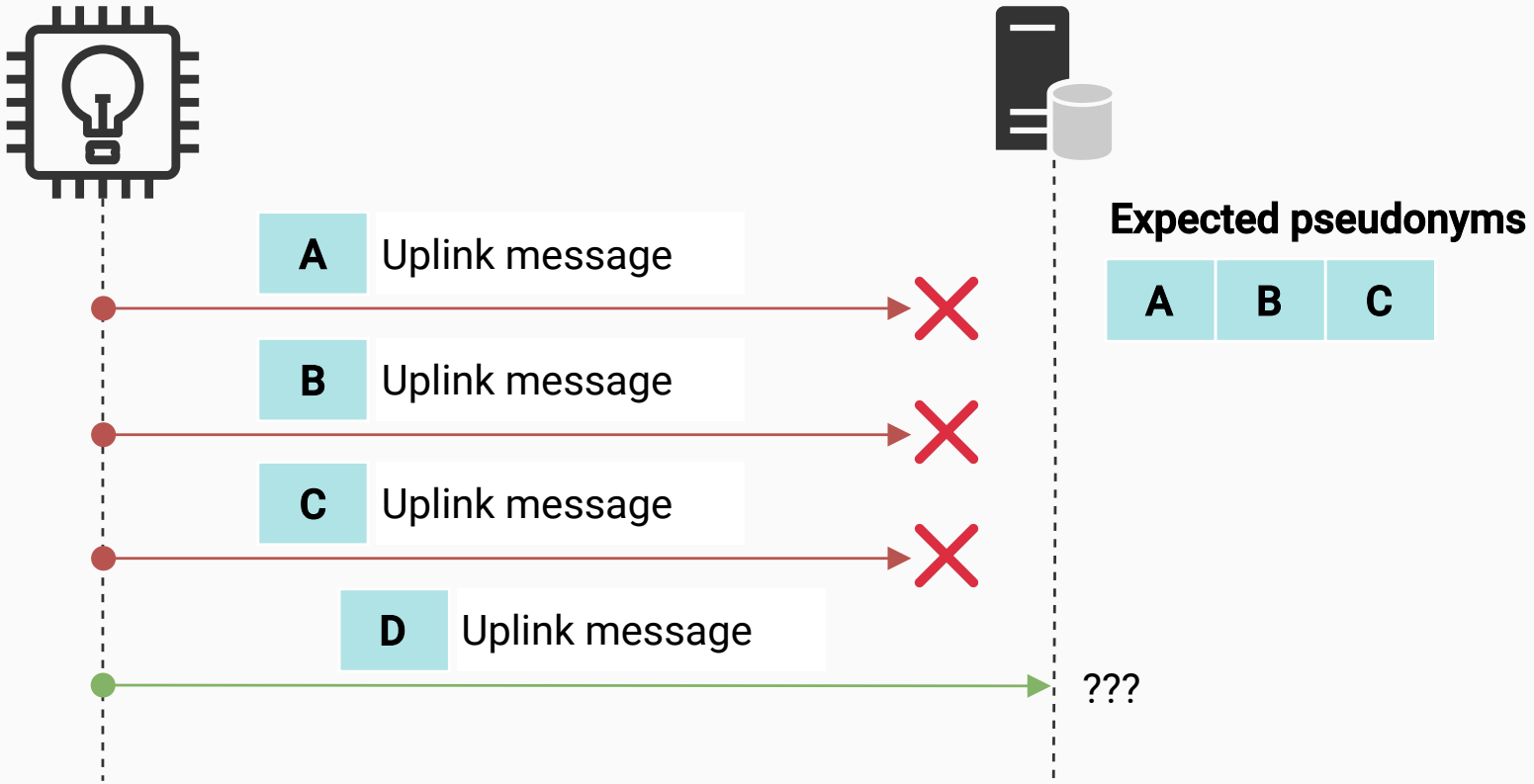
Server-side identification: resolvable pseudonyms have a higher overhead.



For *N* active devices, and *c* collisions.

Evaluation (3): desynchronization

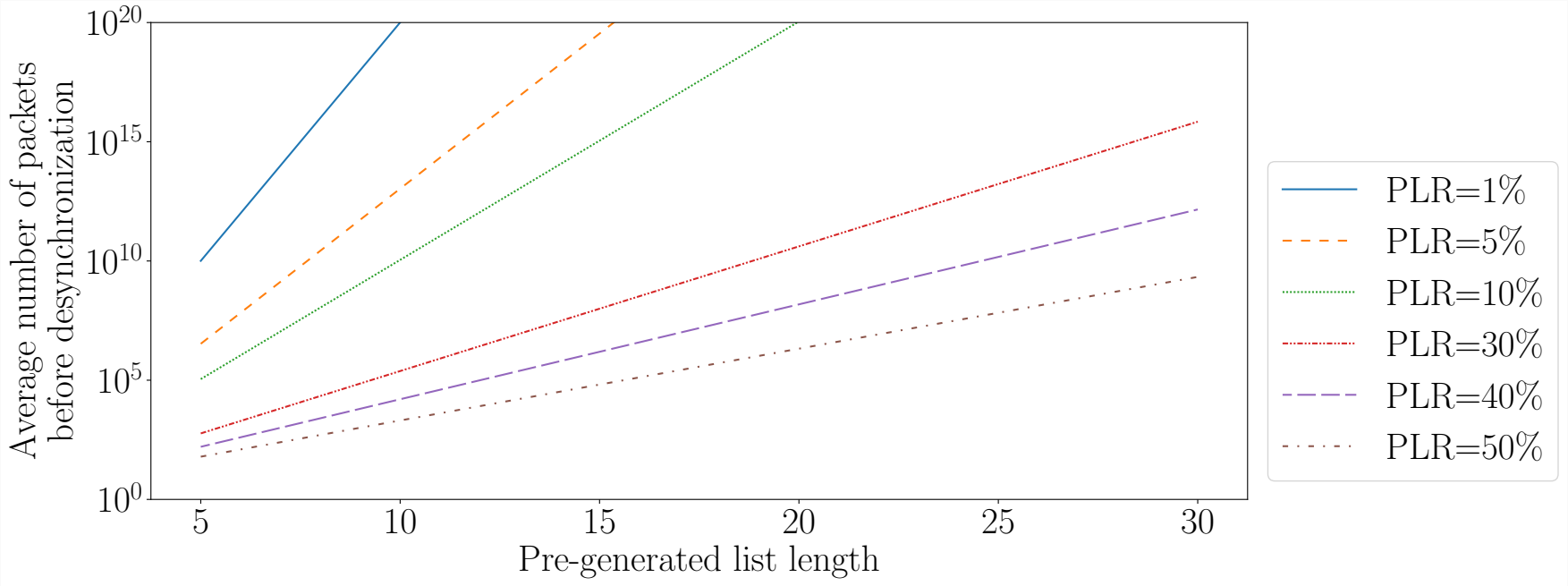
Consecutive packet losses may lead to desynchronization.



For 3 pre-generated pseudonyms.

Evaluation (3): desynchronization

Negligible desynchronization probability observed both **analytically** and empirically.



Devices desynchronize after 11 years, even with 50% PLR (with 1 message per hour and 15 pre-generated pseudonyms).

Evaluation (3): desynchronization

# of pre-generated pseudonyms	5	10	**15**	30
Devices desynchronized at least once	30.5%	9.6%	**4.8%**	0.0%

Empirically, pre-generating enough pseudonyms protects devices from desynchronization.

Evaluation (4): summary

Properties	Resolvable	Sequential
$\mathcal{P}_1$: Unlinkability	✅	✅
$\mathcal{P}_2$: Minimal communication overhead	✅	✅
$\mathcal{P}_3$: Legacy support	✅	✅
$\mathcal{P}_4$: Low computation/memory overhead	❌	✅
$\mathcal{P}_5$: Reliability	✅	✅

Conclusion

Initial observations

- Stable identifiers enable activity inference and tracking.
- Temporary pseudonyms are a solution.
- Existing approaches are not adapted to LoRaWAN.

Experiments

- Adapted and evaluated 2 privacy-preserving pseudonym schemes.

Findings

Sequential pseudonyms are:

- Adaptable to LoRaWAN
- Better than resolvable pseudonyms
  - Lower computation overhead
  - Lower risk of collisions
  - Negligible risk of desynchronization

References I

[1] E. Bäumker, A. Miguel Garcia, and P. Woias.
Minimizing power consumption of LoRa ® and LoRaWAN for low-power wireless sensor nodes.
Journal of Physics: Conference Series, 1407(1), Nov. 2019.

[2] L. Casals, B. Mir, R. Vidal, and C. Gomez.
Modeling the energy performance of LoRaWAN.
Sensors, 17(10):2364, 2017.

[3] P. Ginsparg.
How many coin flips on average does it take to get n consecutive heads?
https://www.cs.cornell.edu/ ginsparg/physics/INFO295/mh.pdf, 2005.

[4] B. Greenstein, D. McCoy, J. Pang, T. Kohno, S. Seshan, and D. Wetherall.
Improving wireless privacy with an identifier-free link layer protocol.
In *MobiSys*, 2008.

[5] Q. Liu, Y. Mu, J. Zhao, J. Feng, and B. Wang.
Characterizing packet loss in city-scale LoRaWAN deployment: Analysis and implications.
In *IFIP*. IEEE, 2020.

References II

[6] K. Tsai, F. Leu, I. You, S. Chang, S. Hu, and H. Park.
Low-power AES data encryption architecture for a LoRaWAN.
IEEE Access, 7, 2019.

Bonus: why resolvable addresses are still better than MIC alone

On average, the cost of re-identification for the Network Server is:

$$C_s + C_f \frac{N-1}{2}$$

For N active devices, C_s the cost for a successful identification, C_f the cost for an unsuccessful identification.

Identification computation costs (in number of AES encryption blocks).

Schemes	C_s	C_f	**Average**
MIC only	4	4	$4 + 4\frac{N-1}{2}$
Resolvable (hash + MIC)	6	2	$6 + 2\frac{N-1}{2}$

Hence, for networks with > than 3 devices, resolvable generate less overhead than directly computing the MIC.

Bonus: LoRaWAN 1.1 key derivation specifications

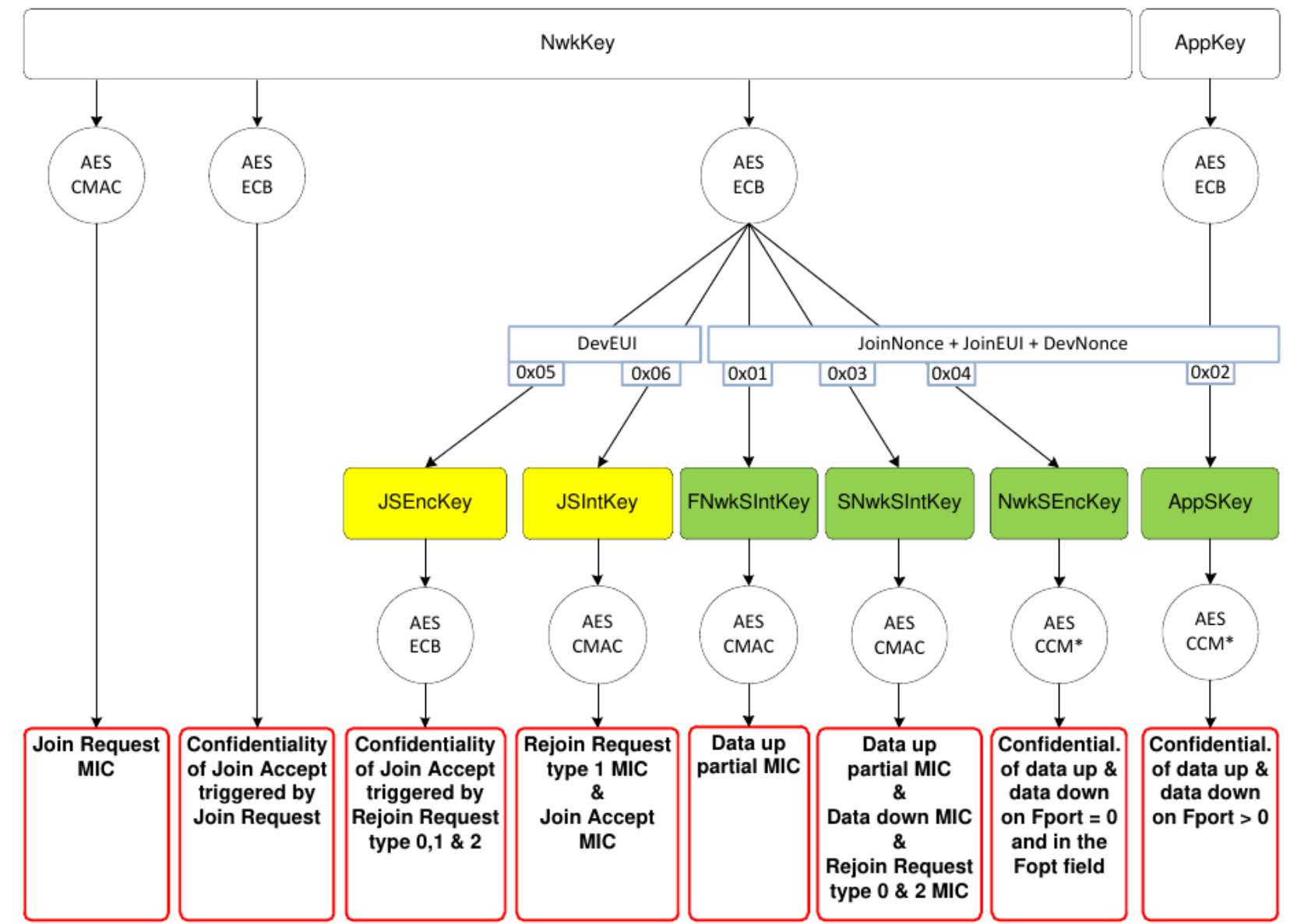


Figure 49 : LoRaWAN1.1 key derivation scheme