future Internet of Things (IoT) connectivity

Petar Popovski
Aalborg University
Denmark
petarp@es.aau.dk

IEEE Future Networks Tutorials @ 6G summit, Levi, Finland, 24-26 March 2019
outline

- future connectivity landscape
- Ultra-Reliable Low Latency Communication
- massive Machine Type Communication (mMTC)
- at the intersection of massive and ultra-reliable
- IoT and distributed ledgers
5G wireless connectivity landscape

enhanced Mobile Broadband

Ultra-Reliable
Low Latency
Communication

data rate

reliability

massive

URLLC

eMBB

mMTC

massive

Machine Type Communication
the connectivity eigenvectors

IA: Industrial Automation
VR: Virtual Reality
V2V: Vehicle-to-Vehicle

IA-1
VR-2
VR-1
V2V-3
future wireless connectivity landscape

5G

but a lot of (great!) other wireless systems

- connectivity type not necessarily provided by the 5G radio interface
- LPWA, 802.11ah, etc.
distilled service requirements

eMBB
- acceleration of 4G, large payloads, active over longer periods
- maximize rate, moderate reliability (e.g. 10E-3)

mMTC
- fix low rate, unknown active subset from a massive device set
- maximize arrival rate, low reliability (e.g. 10E-1)

URLLC
- intermittent transmissions, but from a much smaller device set
- offer high reliability (e.g. 10E-5) while localized in time
IoT use cases

mMTC

- environmental monitoring
- large infrastructures: roads, ports, industrial plants
- parking
- smart agriculture
- management of object fleets: vehicles, bicycles
outline

- future connectivity landscape
- Ultra-Reliable Low Latency Communication
- massive Machine Type Communication (mMTC)
- at the intersection of massive and ultra-reliable
- IoT and distributed ledgers
Ultra-Reliable Low Latency Communication (URLLC)

- use cases and requirements
- communication-theoretic principles of URLLC
- access networking
- two enablers: massive MIMO and multi-connectivity
- statistical aspects and learning for URLLC
URLLC vs. URC (ultra-reliable communication)

early METIS proposal [1]:

- URC over a long term: latency > 10 ms
- URC over a short term: latency ≤ 10 ms

the 5G community has opted to couple the ultra-reliability and low latency into URLLC
- this needs to be revisited

at least two cases for long-term URC

resilient connections with large latency budget

mobile health, remote monitoring
disaster and rescue
two general types of ultra-reliable applications

- cable replacement
  how would we design a system if we could trust to the wireless as much as to the wired?
two general types of ultra-reliable applications

- cable replacement

- "native" wireless applications

  which new systems can we think of once we are empowered with wireless connectivity?
URC/URLLC use cases

- commercial and public safety
- industrial control and automation
- smart energy and smart grid
- digital remote interaction
- V2X and UAV control
- Augmented Reality (AR)

credit: University of Houston
URC/URLLC use cases

potential for revolution in IoT summarized [2]:

smart, connected products offer

exponentially expanding opportunities for

new functionality, far greater reliability,

much higher product utilization, and capabilities that cut across and transcend traditional product boundaries.

ultra-reliable connectivity takes this to the next level as a fundamental design feature

- the best URC/URLLC applications are yet unknown

3GPP requirements for URLLC [3]

- reliability requirement of $1 - 10^5$ (i.e. 99.999 %) with a user-plane radio latency of 1 ms for a single transmission of 32-byte long packet.

- average user-plane radio latency of 0.5 ms for both uplink and downlink, w/o an associated reliability value.

- the most stringent requirements come from factory/process automation, V2X, tactile Internet

- these requirements are incomplete in statistical sense

Ultra-Reliable Low Latency Communication (URLLC)

- use cases and requirements
- communication-theoretic principles of URLLC
- access networking
- two enablers: massive MIMO and multi-connectivity
- statistical aspects and learning for URLLC
a simple communication-theoretic model

\[ y = h \cdot \alpha \cdot x + z + w \]

objective: find \( \alpha \) and, if \( \alpha = 1 \), find also \( x \)
a simple communication-theoretic model

\[ y = h \cdot \alpha \cdot x + z + w \]

ultra-reliability requires to

- model accurately the known unknowns
- bound the impact of the unknown unknowns
- the standard culprit \( z \) seems easy
- interference can be arbitrarily varying
a simple communication-theoretic model

\[ y = h \cdot \alpha \cdot x + z + w \]

sources of uncertainty

- Gaussian noise \( z \) dealt with digital coding schemes
- \( h \) is the problem of channel estimation and channel knowledge
- \( w \) is a matter of interference management and spectrum regulation
  - spectrum license is paid to acquire the right to control interference.
the user activity $\alpha$

the worst case is when there is no prior information about the user activity

- random access

grant-free access means that the packet reception in the uplink is not conditioned on a correct downlink reception

- can improve latency, even reliability

its uncertainty is removed by

- scheduling
- the receiver predicts the activity variable
latency-reliability characterization

one-way vs. two-way latency

reliability = \Pr(\text{latency} \leq t)

1

1-Pe

diversity: time
frequency
antennas
interfaces
design targets

broadband rate-oriented systems

ultra-reliable low latency communication (URLLC)

reliability vs. latency
understanding the tradeoffs

- fix latency to $T$
- fix data size to $D$
  - the minimal required rate is
    \[ R_{\text{bps}} = \frac{D}{T} \text{ [bps]} \]
- fix bandwidth to $B$
  - this gives $2BT$ channel uses.
  - the rate per channel use is
    \[ R_{\text{bps}} = \frac{D}{2BT} = \frac{R}{2B} \text{ [bpcu]} \]
- fix the energy spent in $T$
  - this, along with the channel and noise, leads to the SNR.
- the maximal reliability $1 - \varepsilon$
  is implied by the selection of the previous parameters.
some observations on the model

- low latency \( T \) does not immediately imply short blocklength
  - blocklength \( N = 2BT \) regulated by the bandwidth \( B \)

- relation between activity \( \alpha \) and energy
  - if Bob has a hypothesis that \( \alpha = 0 \), then Alice spends resources to enable detection and Bob spends resources to detect.
  - if Bob has a hypothesis that \( \alpha = 1 \), then he spends a lot of energy listening all the time.
an example of a short packet format

UNB (ultra narrowband) system

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>preamble</td>
<td>frame sync.</td>
<td>4 bytes</td>
</tr>
<tr>
<td>end-device ID</td>
<td></td>
<td>4 bytes</td>
</tr>
<tr>
<td>payload</td>
<td></td>
<td>0..12 bytes</td>
</tr>
<tr>
<td>authentication</td>
<td></td>
<td>var.</td>
</tr>
<tr>
<td>frame check sequence</td>
<td></td>
<td>2 bytes</td>
</tr>
</tbody>
</table>

reliability of the packet reception is a **product** of the reliabilities of different parts

\[
Pr(\text{success}) = Pr(\text{PA}) \times Pr(\text{sync}) \times Pr(\text{ID}) \times Pr(\text{data}) \ldots
\]
communication theory and protocol information
at short blocklengths there is a penalty that keeps the rate away from capacity.

fundamental theory of finite blocklength transmission
illustration of the finite blocklength effect

- 16 bytes control information
- 16 bytes data
- reliability 99.999%
- latency 1 ms
- reference SNR for 100 kHz

\[ \gamma = \gamma_0 \frac{B_0}{B} = \gamma_0 \frac{2B_0T}{N} \]
illustration of the finite blocklength effect

\[ B = \frac{1}{2} \frac{N}{\gamma} \log_2 \left( 1 + \frac{\gamma}{N} \right) \]

where \( B \) denotes the channel capacity and dispersion, respectively. Note that the SNR also depends on the bandwidth \( \gamma \) and the number of available channel uses \( N \).

We consider two cases:

1) Data and metadata are encoded jointly, the probability of correct packet reception is

\[ \epsilon = \frac{1}{2} \left( 1 - \sin \left( \frac{\pi}{2} \log_2 \left( 1 + \frac{\gamma}{N} \right) \right) \right) \]

2) Data and metadata are encoded separately and the probability of correct packet reception is

\[ \epsilon = \frac{1}{2} \left( 1 - \sin \left( \frac{\pi}{2} \log_2 \left( 1 + \frac{\gamma}{N} \right) \right) \right) \]

The trick with increasing the blocklength in order to attain good performance requires a high reference SNR. Nevertheless, the same trick of increasing the blocklength in order to attain good performance requires a high reference SNR.
frequency diversity for URLLC

two cases dependent on the relation between the bandwidth $B$ and the coherence bandwidth $B_c$

- $B \leq B_c$ gain in reliability comes from increasing the blocklength

- $B > B_c$ this is "true" frequency diversity, bringing independently faded channel uses.
mixing data and control information has energy cost

the notion of frame in cellular systems should be revisited
connection between packetization and energy efficiency

separated data and metadata
useful for energy efficiency
data for Bob, Carol turns off her receiver after the metadata

joint data and metadata
better coding of the metadata
however, everybody decodes everything
example [KT17]:
a theoretical treatment of downlink framing

downlink communication to K users

- a user is active and there is a packet for her with probability q.
- the message for each active user is drawn randomly from a set of predefined message sizes.
- metadata should inform about
  - who is active
  - the message size.

example: a theoretical treatment of downlink framing

conventional framing with pointers

alternative framing
new tradeoff arises for short packets
- latency is minimized when all packets are jointly encoded;
- power is minimized when each packet is encoded separately.

K=32 users
outline

- Ultra-Reliable Low Latency Communication (URLLC)
  - use cases and requirements
  - communication-theoretic principles of URLLC
  - access networking
    - two enablers: massive MIMO and multi-connectivity
    - statistical aspects and learning for URLLC
what should access networking solve

- resolve uncertainty in user activity $\alpha$
- run the auxiliary operations, notably synchronization
- decode metadata and data
- interact with the higher-layer protocols
deterministic arrivals

- typical in closed-loop isochronous communications
  - motion control, factory automation

![Diagram showing Factory Automation and Motion Control with different real-time classes and performance requirements]

figure from

deterministic arrivals

- the value of $\alpha$ is known a priori
- reliability critically dependent on time sync
- error probability

$$\epsilon_{\text{det}} = 1 - (1 - \epsilon_{\text{sync}})(1 - \epsilon_D)(1 - \epsilon_A)$$

- data
- ack
stochastic arrivals

four-step access

\[ \epsilon_{\text{sto4}} = 1 - (1 - \epsilon_{\text{sync}})(1 - \epsilon_R)(1 - \epsilon_G)(1 - \epsilon_D)(1 - \epsilon_A) \]

- random access affects the request
- makes sense if activity is (very) intermittent
stochastic arrivals

practical protocols involve more steps
stochastic arrivals

three-step access
- has the issue of resource waste

\[
\epsilon_{sto3} = 1 - (1 - \epsilon_{sync})(1 - \epsilon_G)(1 - \epsilon_D)(1 - \epsilon_A)
\]

grant-free access
- minus: high error rate for the data
- plus: low latency, decreased overhead

\[
\epsilon_{sto2} = 1 - (1 - \epsilon_{sync})(1 - \epsilon_D)(1 - \epsilon_A)
\]
Ultra-Reliable Low Latency Communication (URLLC)

- use cases and requirements
- communication-theoretic principles of URLLC
- access networking
- two enablers: massive MIMO and multi-connectivity
- statistical aspects and learning for URLLC
massive MIMO for URLLC

factory scenario
massive MIMO and ultra-reliability

pros
- very high SNR links
- quasi-deterministic links, fading immunity
- extreme spatial multiplexing capability

cons
- expensive CSI acquisition procedure
- additional protocol steps
massive MIMO and ultra-reliability

- downlink, single-antenna terminals
- use the channel structure and the propagation path directions
- cluster-based propagation model
- channel of the k-th terminal

\[
H^{(k)} = \sum_{i=1}^{N_P^{(k)}} \alpha_{i}^{(k)} s_{i,rx}^{(k)} s_{i,tx}^{(k)} H_{i,rx,i,tx}^{(k)}
\]

#paths

fast fading

channel structure
massive MIMO and ultra-reliability

- channel structure used for zero-forcing beamforming
- inter-terminal interference removed by selection of spatial paths
- intra-terminal combining of paths

\[ F_{ZF}^{(1)} = P_{\perp} V^{(2)} \]

inter-terminal precoding

linear combining

singular vectors

\[ \bar{V}^{(1)} \]

\[ w \]
terminals with a N=1 antenna

For the selected simulation parameters, the following observations can be highlighted:

- The general tendency is that performance gets better with an increased exploitation level about the channel at the transmitter.

- There is a notable exception when the terminals are equipped with multiple antennas and receive diversity is exploited. In Fig. 8, for \( N=4 \), the non-coherent strategy ("All SV - NCoh") performs the best. Hence, from a BER perspective, it is preferable to transmit the signal in a non-coherent fashion along each singular vector. The non-coherent transmission is compensate for by a receive coherent processing by multiple antennas that allows the extraction of diversity.

- Depending on the level of CSI exploited at the transmitter, space multiplexing is not always favourable.

VI. MULTI-CONNECTIVITY AND INTERFACE DIVERSITY

The mobile devices today have multiple radio interfaces and it is likely that many of the future devices will have that as well. This is also an indicator that the 5G radio interfaces will be deployed along with other radio interfaces. From the perspective of URLLC, the existence of multiple interfaces offers an additional degree of diversity that can be used to fulfil the stringent latency-reliability requirements. This is commonly known as multi-connectivity [47], while here we use the terms link diversity or interface diversity in order to emphasize the diversity role played by the availability of multiple different communication interfaces.

The idea of using multiple links or interfaces simultaneously is fairly natural and it has already emerged in some settings. In the context of 3GPP systems, LTE has supported Multi-Connectivity through Carrier Aggregation (CA) and Dual Connectivity (DC) since rel. 10 and 12, respectively. However, in this case the objective is throughput enhancement. Recently, in Rel. 15, Packet Duplication was introduced by 3GPP to boost reliability [47]. The data packet is duplicated on PDCP and transmitted on independent channels, either from the same eNB on different carriers via CA or from different eNBs using DC.

Packet Duplication in Multi-Connectivity architectures are excellent for mitigating losses due to fading and interference on individual links or temporary scarcity of air interface resources. Nevertheless, the reliability of the end-to-end connectivity relies on the correct functioning of an infrastructure and core network, often belonging to a single operator. While infrastructure and core networks are based on redundant solutions, they are still subject to single Point-of-Failure (PoF), e.g. through equipment misconfiguration. This reliance on a single network infrastructure can be mitigated by providing diversity not only at a link level, but also at a level of a communication interface or a path, as illustrated in Fig. 9.

The concept of Interface Diversity (IFD) was studied in [48]. Interface Diversity provides an independent path from the UE to the internet (cloud), by the use of a different wireless technology and/or a different mobile network operator. That is,
terminals with a N=4 antennas

solid: spatially multiplexed

dashed: time multiplexed
multi-connectivity and interface diversity

A. Reliability Model and Numerical Illustration

Let us initially consider the two-link/interface instances of data segmentation and packet-level coding. The key benefit of IFD is that there is no dependency on a single point of failure in the access network part. However, IFD requires that source and destination devices are configured to duplicate packets and handle multiple received copies.

The relations represented by (18) are for a single link, whereas the IFD are for for a single link, respectively.

In comparison, consider the plot in Fig. 11, where the mobile host outage is instead dominated by the core outage probability. Further, the plot reveals that DC is better than using a single link, unless when the link outage is very low, and the end-to-end outage is instead dominated by the core outage probability.

Dual Connectivity

Interface Diversity
parallel use of multiple IoT interfaces

parallel use of multiple interfaces suitable to support the latency-reliability characteristics

combination through repetition or more sophisticated coding.
multi-connectivity and interface diversity

compared to dual connectivity, interface diversity avoids a single point of failure.

various coding schemes possible

interface diversity
experimental results

results based on lab measurements

- 1 day, 100 ms interval
- Wi-Fi
  - achieves 10 ms for 90% of packets
  - but 99% requires almost 100 ms
- cellular: LTE and HSPA
  - also requires ~100 ms for 99%
- duplication (1 copy per IF)
  - 99% at 25 ms
  - 99.999% at 60 ms latency
Outline

- Ultra-Reliable Low Latency Communication (URLLC)
  - use cases and requirements
  - communication-theoretic principles of URLLC
  - access networking
  - two enablers: massive MIMO and multi-connectivity
  - statistical aspects and learning for URLLC
- Ultra-Reliable Low Latency Communication (URLLC)
  - use cases and requirements
  - communication-theoretic principles of URLLC
  - access networking
  - two enablers: massive MIMO and multi-connectivity
  - statistical aspects and learning for URLLC
wireless channel model behavior in ultra-reliable regime

currently there is lack of experimental evidence for URC-relevant statistics of wireless channels

initial analysis of common wireless channel models in URC regime

- block fading

- $P_R$ is the minimal SNR to decode data rate $R$

- the analysis reveals the URC-behavior:

$$
\Pr \left( P_R \frac{R}{P} < L \right) \approx \varepsilon \approx \alpha \left( \frac{P_R}{P} \right)^{\beta}
$$
two-wave model with equal amplitudes represents one of the worst cases
indoor physical setup

indoor case dominated by diffuse components, good for high reliability

Alice sells ultra-reliable wireless to Bob

- how does Alice measure the reliability performance?
- under what conditions is ultra-reliability guaranteed?
Bob poses inconvenient questions and wants to be convinced.

- what is the true distribution $F$ used to calculate the reliability?
- what if the true channel distribution differs from $F$?
- what if we have no knowledge of $F$ at all?
back to modeling sources of uncertainty

a communication engineer models known unknowns

\[ y = \alpha \cdot h \cdot x + z + i \]

URLLC requires to bound the impact of the unknown unknowns in the models

- interference in unlicensed spectrum
  is almost unknown unknown
a simple error model

\[
Pr(E) = Pr \left( \frac{\gamma_S}{1 + \gamma_I} < \gamma_{th} \right)
\]

- in absence of interference, we need to characterize the lower tail of \( \gamma_S \)
- if \( \gamma_S \) is known, we need to characterize the upper tail of \( \gamma_I \)
channel uncertainty in ultra-reliability

- assume that the interference is absent.
- we (somehow) know that the channel is Rayleigh.
- the target error rate is $\varepsilon_U$, average SNR is $\bar{\gamma}_S$

how do we choose the rate $R$?

\[
Pr(E) = Pr(\log_2 (1 + \gamma_S) < R)
\]

\[
R = \log_2 \left( 1 + \bar{\gamma}_S \ln \left( \frac{1}{1 - \varepsilon_U} \right) \right)
\]

where is the problem?
statistical machine learning in a nano-shell

- we are observing training data of $n$ values $X^n$ sampled from an unknown distribution $F(x)$

- we use the training sample in order to predict some function of the future values sampled from the same distribution

- the distribution $F_\theta(x)$ may be unknown up to a parameter $\theta$.

- how to choose the sample size in order to satisfy some requirements of correctness when predictions are made
statistical framework for ultra-reliability

three key elements

- **model selection**
  - parametric models for $F$
  - non-parametric estimation of $F$

- **learning**
  - generate an estimate $\hat{F}$ using a training sample $X^n$
  - bad training data leads to a bad estimate $\hat{F}$

- **performance evaluation**
  - specification of the packet error rate only is **insufficient**
performance evaluation done right

learning can affect various decisions the system
- rate selection is the simplest example
- other examples:
  choice of resources in slicing, user admission, etc.

we define two types of reliability

Averaged Reliability (AR)

Probably Correct Reliability (PCR)
Averaged Reliability (AR)

- calculated as a reliability that is averaged across all possible training samples \( \{X^n\} \)

\[
\sup_{F \in \mathcal{F}} \overline{\text{PER}} \leq \epsilon
\]

- suitable for dynamic environments, where separate training is not feasible
Probably Correct Reliability (PCR)

- calculated as a reliability that obtainable for a specific training sample $X^n$

$$P(\text{PER} > \epsilon) \leq \xi$$

- suitable for a static environment where training precedes operation
**parametric vs. non-parametric**

- if parametric estimation is used then we need to have a good guess about the channel model
  - mismatch possible: channel assumed Rayleigh, but is actually Nakagami

- non-parametric estimation does not assume any distribution
  - number of required training samples overwhelming, in the order of $1/\varepsilon$, which can go beyond $10^9$!
example performance evaluation

- average rate achieved through learning normalized by the average rate achievable if the true distribution is known

![Graphs showing parametric vs non-parametric rate-selection under Rayleigh fading with different values of \( \epsilon \) and \( \xi \).](image)

(a) \( \epsilon = 10^{-3}, \xi = 10^{-3} \)

(b) \( \epsilon = 10^{-4}, \xi = 10^{-3} \)

(c) \( \epsilon = 10^{-5}, \xi = 10^{-3} \)
using machine learning beyond rate selection

Communication layer

Alice

- Data model and learning
- Data source/destination
- Protocol information

TXRX1

Bob

- Data model and learning
- Data source/destination
- Protocol information

Noise source

Data analytics/machine learning layer
example: creating side information for protocols

- an immediate approach: remove some steps (e.g. grant free)

- ML-based approach:
  - predict the packets and their content
  - use that as side information for decoding

\[ P_S = P_S(grantRQ)P_S(grant)P_S(data)P_S(ACK) \]
summary and outlook

- ultra-reliable wireless has the potential to profoundly change systems and devices

- essential:
  - short packet transmission
  - communication-theoretic attention to the control information
  - every step in the protocol needs a careful reliability design
  - careful use of diversity

- large number of steps in real protocols impair reliability and latency
  - lean protocol design
summary and outlook

- ultra-reliability needs a firm statistical basis, both in system design and system verification
  - risk-based methods

- we have defined two performance measures that generalize the simple PER reliability demand

- shown the utility of the statistical learning approach in the simple rate selection scenario

- but more involved communication settings can also embrace the approach
outline

- future connectivity landscape
- Ultra-Reliable Low Latency Communication
- massive Machine Type Communication (mMTC)
- at the intersection of massive and ultra-reliable
- IoT and distributed ledgers
massive Machine Type Communication (mMTC)

- use cases and requirements
- access problem and architectures for massive access
- communication models
massive MTC as a key IoT enabler

- anything that can be connected, will be connected
  - but also everything becomes vulnerable

- heterogeneity in terms of
  - device types
  - applications
  - traffic patterns
  - performance requirements
MTC/IoT example: water metering

important problem to be solved: water leakage

From: http://www.smartm2mpt.pt/Pages/EN/Services/WaterTelemetry.aspx
MTC/IoT example: vending machine

important problem to be solved: food waste

From: http://www.smartm2mpt.pt/Pages/EN/Services/VendingMachineSupervision.aspx
mMTC use cases

- environmental monitoring of large areas
- large infrastructures: roads, ports, industrial plants
- smart metering
- management of object fleets: vehicles, bicycles
requirements for mMTC

- 10-15 years device battery life
- extended coverage
- 300,000 connected devices per cell
  - New Radio (NR) goes to 1,000,000 devices/km²
- low complexity
- efficient transmission of sporadic small payloads
- per-packet reliability relatively low, but
  - some stringent reliability constraint over an extended period (e.g. one packet per day)
  - joint reliability requirements put to a group of IoT devices.
outline

- massive Machine Type Communication (mMTC)
  - use cases and requirements
  - access problem and architectures for massive access
  - communication models
canonical access problem with a single BS

the set of nodes that want to send to the BS in the uplink is unknown
The access problem in mMTC

Quick assessment of the overhead

- \( N \) nodes, \( K \) active
- If \( N, K \) are known, then the uncertainty is
  \[
  \log_2 \left( \frac{N}{K} \right) \text{ [bits]}
  \]

- Maximized when \( K = \frac{N}{2} \)
- However, mMTC works with assumption of sporadic, low-intensity access with \( K \ll N \), pointing to random access protocols
cellular with capillary access network

pros

- remove the contention from the BS and put it to the aggregators
- aggregated packets more efficient w.r.t. overhead
- support low-power operation
- additional infrastructure cost and latency
cellular with capillary access network

- **cons**
  - moves the contention to unlicensed spectrum, less predictable quality of access
  - in the *extreme case* it boils down to the same access problem.
- efficient random access aggregator-BS due to larger aggregated packets

cellular with capillary access network
Remote Radio Heads (RRHs) with limited functionality
massive data access supported by collaborating through fronthauls

cloud architecture

- loss due to digital transport over the fronthaul
- local decision (edge computing) vs. centralized decision
- edge computing one of the enablers of low latency
the access problems in vehicular setting

all-to-all broadcast

- the objective is to spread the message to all neighboring vehicles
- problem of half-duplex: cannot receive message while sending

the access problems in vehicular setting

network model
periodic vehicular broadcast

tradeoff between
transmission to be heard
listen to hear the others

A transmits in shaded interval
patterned packets erased due to noise
yet another access problem in vehicular setting

distributed handshake

- three-way handshake between each pair of vehicles
- besides broadcasting, each device need to infer which other device has received its data

outline

- massive Machine Type Communication (mMTC)
  - use cases and requirements
  - access problem and architectures for massive access
  - communication models
communication model

- multiple access channel
- star topology
  - single hop
  - common receiver  
    (base station, access point)

- the full extent of the problem:
  - No synchronization
  - No CSI
  - The base station does not know:
    • Who is active
    • How many are active, i.e., the value of $K$
  - $N$ (or some upper bound on its value)  
    is assumed to be known
  - Active users want to send data to the BS
  - Active users lack coordination
  - $K \ll N$ – channel partitioning/taking  
    turns protocols become highly inefficient

$K$ active users  
$N - K$ inactive users
communication model

this is a slotted model

single BS received signal in the $k$ –th slot:

$$y_k = \sum_{n=1}^{N} h_{kn} a_{kn} x_{kn} + z_k$$

$N$ - total number of users;

$h_{kn}$ - wireless channel coefficient during the $k$ –th slot;

$a_{kn}$ - the activity of $n$–th user during the $k$ –th slot,

$$a_{kn} = 1 \text{ if the user is active and } a_{kn} = 0 \text{ otherwise;}$$

$x_{kn}$ - packet/symbol sent by the $n$–th user during the $k$ –th slot;

$z_k$ - noise in the $k$ –th slot.
communication model

the same model used for
- reception of a symbol
- reception of a packet over a slotted quasi-static channel

the important point is that the channel is not changed during the single transmission

we will deal with models in which the channel is not changed over multiple transmission slots/symbols
sources of uncertainty

\[ y_k = \sum_{n=1}^{N} h_{kn} a_{kn} x_{kn} + z_k \]

all of the following:

\[ h_{kn}, a_{kn}, x_{kn}, z_k \]

the central role in mMTC is played by the uncertainty contained in the activity variable \( a_{kn} \):

- sometimes the total number of active users \( \sum_{n=1}^{N} a_{kn} \)
  can be assumed known or it can be estimated

for modeling purpose we can make \( a_{kn} \):

- part of the channel \( h'_{kn} = h_{kn} a_{kn} \), leading e.g.
  to Bernoulli-Gauss distribution for Rayleigh channel
- part of the transmitted symbol \( x'_{kn} = x_{kn} a_{kn} \), leading to having a special "empty" symbol.
collision model

- neglect noise
- packet is the atomic unit, not symbol
- put all channel coefficients $h_{kn} = 1$

justified through channel inversion:
- the BS sends broadcast
- each device $n$ estimates its $h_{kn}$ due to reciprocity
- if not too much transmit power is required, i.e. $\frac{1}{\|h_{kn}\|^2} \leq P_{\text{max}}$
  then if $a_{kn} = 1$ the device transmits $\frac{x_{kn}}{h_{kn}}$
- if $\frac{1}{\|h_{kn}\|^2} > P_{\text{max}}$ then the device sets $a_{kn} = 0$
collision model

\[ y_k = \sum_{n=1}^{N} a_{kn} x_{kn} \]

here \( x_{kn} \) is a fixed-rate signal, decodable only if there is no interference

this brings us to the classical objective of random access:

choose the probability of being active in order to maximize the probability that a single node is active in a slot.

- Q: how to choose then the probability of \( a_{kn} = 1 \)?
collision model with capture and SIC

\[ y_k = \sum_{n=1}^{N} h_{kn}a_{kn}x_{kn} + z_k \]

here we do not assume that the channel coefficients are 1 but they are known or can be estimated.

fixed-rate signal that can be decoded when the SIR condition is satisfied

- strongest user attempted to be decoded first
- if successful, cancelled and procedure repeated.
standard information-theoretic model and capacity

\[ y_k = \sum_{n=1}^{N} h_{kn}x_{kn} + z_k \]

here the uncertainty in \( a_{kn} \) does not exist

- either \( a_{kn} = 1 \) for all \( n \)
- or all devices, including the BS, know which \( a_{kn} = 1 \)

the objective is to find \( n \)-tuples of achievable rates
assuming very large (infinite) codewords
relation between capacity and collisions

- interpreting a collision through a multiple access region
- role of successive interference cancellation
vector communication model

vector observed over M slots

\[ y = Sx' + z \]

\[ x' = [h_1 a_1 x_1 \ h_2 a_2 x_2 \ h_3 a_3 x_3 \ ... \ h_n a_n x_n]^T \]

here \( S \) is the \textit{signature matrix}

- has the role of a spreading code in a CDMA model
- the model shows relation to compressed sensing
- signatures can operate on packet level as protocol sequences
**$K$-out-of-$N$-user MAC**

- also known as partial-activity model
- the users contend with signatures
  - every user has a unique signature!
- BS wants to learn who is active
  - but not through random access
- the signatures are constructed such that they can be reconstructed from an observation that contains any combination of no more than $K$ of them.
**$K$-out-of-$N$-user MAC**

- **the number of contending users** $K'$ is a random variable
  - we actually “hope” that $K' \leq K$
  - generalizes collision, by adopting $K > 1$

- **several variants:**
  - $K$-out-of-$N$-user Gaussian MAC
  - $K$-out-of-$N$-user OR MAC
  - $K$-out-of-$N$-user integer-adder MAC
  - $K$-out-of-$N$-user real-adder MAC
  - ....

![Diagram](image.png)

- base station
- signatures
- no more than $K$ active users
**K-out-of-N-user binary-input integer-adder MAC**

- a simple, but insightful model
  - noise is negligible (via application of, e.g., channel code)
  - powers at the point of reception are equal

![Diagram of K-out-of-N-user binary-input integer-adder MAC](image)
$K$-out-of-$N$-user binary-input integer-adder MAC

- The channel output is:
  \[ y = \sum_{\ell \in \mathcal{L}} s_\ell \]
  where $s_\ell \in \{0,1\}^L$ and $y \in \mathbb{N}^L$

- The signatures are constructed such that any sum of no more than $K$ of them (i.e., $|\mathcal{L}| \leq K$) can be decoupled into constituent signatures.

No more than $K$ active users
**$K$-out-of-$N$-user binary-input integer-adder MAC**

- The best known construction in terms of minimizing signature length $L$ is the one by Lindström (1975):
  \[ L \approx K \log N + 1 \]
  where $N$ is the total number of users.

- If $K' > K$ (i.e., more than $K$ users active), the decoder declares an error.
  - There is an indicator bit included in the signature – sum of the indicator bits tells the value of $K'$.

---

The asymptotic bound on the minimal signature length $L$ that achieves $K$-out-of-$M$ signature decoding is:

$$\left\lfloor \frac{2K}{\log K} \log M \right\rfloor + 1 \leq L \leq \left\lfloor \frac{4K}{\log K} \log M \right\rfloor + 1$$

many-access channel (MnAC)

ew information-theoretic model suited for mMTC

\[ Y = \sum_{k=1}^{\ell_n} S_k(w_k) + Z \]

- each user accesses the channel independently with \( \alpha_n \)
- if not accessing, a special zero codeword is sent
- if accessing, one of the M messages is sent
- number of possible users \( l_n \) tied to the packet blocklength \( n \)

many-access channel (MnAC)

- Gaussian many-access channel:
  \[ y = \sum_{l=1}^{N_n} s_l(w_l) + z \]

  where \( w_l \) is the message of user \( l \) and \( s_l \in \mathbb{R}^n \)

- If user \( l \) is inactive, then \( w_l = 0 \) and \( s_l(w_l) = 0 \)

- Otherwise, if user \( l \) is active, then:
  - \( w_l \in \{1, ..., D\}, \forall l \),
  - all non-zero message are equiprobable
  - \( s_l(w_l) \) is selected from a codebook of user \( l \), where the codebook is unique
    - i.e., \( s_l(w_l) \) carries the information about the user and the message

The average number of active users is \( K_n = \alpha_n N_n \)
many-access channel (MnAC): achievability

1. If $N_n$ is bounded, i.e., $N_n = N \leq \infty$ for large enough $n$, then

$$R_u = \frac{1}{2N} \log(1 + N P)$$

2. If $N_n$ is unbounded and:
   - $K_n = O(n)$, and
   - $N_n e^{-\delta K_n} \to 0$, and
   - $\alpha_n \to \alpha \in [0,1]$, and
   - $N_n H(\alpha_n) < \frac{n}{2} \log(1 + K_n P)$,

then:

$$R_u = \frac{1}{2K_n} \log(1 + K_n P) - \frac{N_n H(\alpha_n)}{n K_n}$$
many-access channel (MnAC)

- detecting active users of central importance
- related to the problem of sparse recovery (compressed sensing)
- analyzed when both number of users and blocklength go to infinity
- key element is user detection based on signatures
- error defined based on joint correct detection of all users

- the message length grows with n

\[ \ell_n = n \]
\[ \ell_n = n^{1.5} \]
\[ \ell_n = n^2 \]
\[ \ell_n = n^3 \]
many-access channel (MnAC): achievability

\[ R_u = \frac{1}{2K_n} \log(1 + K_n P) - \frac{N_n H(\alpha_n)}{n K_n} \]

- it can be shown that if \( N_n \) grows linearly with \( n \), the penalty term tends to zero
- the strategy that asymptotically achieves the above result consists of concatenation of user signatures and codewords that only represent messages, where the signature length is \( \frac{N_n H(\alpha_n)}{K_n} \)
- for all other cases related to scaling of \( N_n \) and \( K_n \) with blocklength \( n \):
  \[ R_u \to 0 \]
massive grant-free access without user identification

- Gaussian MAC with $h_{kn} = 1$.
- $K$ active users
- $D$ non-zero messages
- all users have one and the same codebook with $D$ codewords
  - No user identification possible!
- an active user chooses its message uniformly at random, independently of any other user
- decoding is done up to a permutation of transmitted messages

massive grant-free access without user identification

- error defined from the perspective of a user
- error occurs if the message is not decoded correctly or more than one user sends the same message
- finite blocklength limiting when the number of users is low, multi-user interference when the number of users is high.

non-coherent channel model

\[ y_k = \sum_{n=1}^{N} h_{kn} a_{kn} x_{kn} + z_k \]

this is the ”most honest” model for random access where both the channel and the user activity are unknown

- approach: make \( x'_{kn} = x_{kn} a_{kn} \) into a single unknown
- codebooks of users chosen in a way to create linear subspaces that can be used to both
  - detect the subset of active users;
  - decode their codewords;

cellular access protocols

- all (mobile) cellular standards use a similar algorithm for the initial connection establishment
- access reservation protocol:

![Diagram showing the access reservation protocol]

- Mobile subscriber
  - Random access identifier
  - Resource request
- Base station
  - Random access response
  - Resource assignment

**Slotted ALOHA based**
slotted ALOHA


- \(N\) users, single access point (AP)
  - homogenous population
  - equal length packets

- random access
  - distributed, decentralized
  - users behave in the same way

- link time is divided in slots of equal duration
  - slot-synchronization of the users is assumed

- users contend for the access to the base station
slotted ALOHA

- each user contends (transmits) with a predefined probability $p_A$
- collision model in which a slot can be
  - Idle
  - Singleton (i.e., containing single transmission)
  - Collision (containing multiple transmissions)
- collision channel model
- feedback after every slot
  - unsuccessful users contend in the next slot with (possibly changed) $p_A$
slotted ALOHA

- **throughput:**
  - measure of efficiency of use of system resources (slots)
  - average fraction of slots with successful transmission attempts
    - i.e., fraction of singleton slots

- **probability that a slot is a singleton:**
  - \( T = \binom{N}{1}p_A(1 - p)^{N-1} \approx Np_Ae^{-Np_A} \)
  - \( T_{\text{max}} = \frac{1}{e} \approx 0.37 \) (when \( Np_A = 1 \))
framed slotted ALOHA


- related to our vector comm model

- each user transmits just once in a randomly selected slot of the frame

- feedback on a frame basis

- throughput:
  - Probability that a slot is a singleton:
    \[ T = \binom{N}{1} \left( \frac{1}{M} \right) \left( 1 - \frac{1}{M} \right)^{N-1} = \frac{N}{M} e^{-\frac{N}{M}} \]
  - \( T_{\text{max}} = \frac{1}{e} \approx 0.37 \) (when \( \frac{N}{M} = 1 \))
collision model with SIC: contention resolution diversity slotted ALOHA


- users repeat their transmission in several randomly chosen slots of the frame
  - same number of replicas per user

- collisions can be exploited!
  - successive interference cancellation (SIC)
  - improves throughput
  - $T \approx 0.55$ for CRDSA with two repetitions per user
SIC in slotted ALOHA-based protocols

- Each successfully decoded replica enables canceling (removal) of other replicas
  - Channel $h$ assumed constant, estimated while decoding the singleton and then used for interference cancellation.

- In the first approximation, it is assumed that interference cancellation is perfect.
irregular repetition slotted ALOHA (IRSA)


- **generalization of CRDSA**
  - No. of replicas can vary across users
  - every user selects its no. of replicas according to a predefined distribution

- **interference cancellation (IC)** can be seen as iterative procedure performed on the graph
  - resembles decoding procedure of erasure correcting codes

- **IC in frame-slotted ALOHA** – analogous to block error-correction coding
successive interference cancellation
Irregular repetition slotted ALOHA

The presented results are for the user degree distribution optimized for vanishingly small packet loss probability in asymptotic scenario.

Logical load $G = \frac{N}{M}$

Physical load is $\Lambda'(1)G = E[|u|] = 3.6G$!

$$R = \frac{1}{E[|u|]} = \frac{1}{3.6} \approx 0.28$$

frameless ALOHA


- idea: Apply paradigm of rateless codes to slotted ALOHA:
  - no predefined frame length
  - slots are successively added until a criterion related to performance parameters of the scheme is satisfied
  - optimization of the slot-access probability and termination criterion
and-or tree evaluation: asymptotic analysis tool

and-or tree evaluation: performance parameters

- and-or tree evaluation shows the expected asymptotic performance based on the statistical graph description expressed through degree distributions $\lambda(x)$ and $\omega(x)$

- probability of user resolution:
  $$P_R = 1 - \lim_{i \to \infty} q(i)$$
  - with the initial value $q(0) = 1$

- asymptotic throughput is:
  $$T = \frac{P_R N}{M}$$
maximal throughput
optimal slot degree (maximal throughput)
probability of user resolution
frameless ALOHA: optimizing the stopping criterion


- single feedback used after $M$-th slot
  - $M$ not defined in advance
  - Analogous to rateless coding framework!

- when to send feedback?
  - E.g., when the throughput is high enough (ideally the highest possible)
frameless ALOHA: optimizing the stopping criterion

- the graph shows the evolution of the probability of user resolution $P_R$ and the throughput $T$ in the asymptotic settings for the optimal $\beta \approx 3.1$
- asymptotically optimal way to maximize throughput:
  - end the contention when the throughput starts to drop
frameless ALOHA: optimizing the stopping criterion

- an example of a typical run of frameless ALOHA in terms of:
  - fraction of resolved users
    \[ F_R = \frac{N_R}{N} \]
  - instantaneous throughput
    \[ T_R = \frac{N_R}{M} \]

heuristic stopping criterion: stop when \( T \) is maximal

genie-aided stopping criterion: fraction of resolved users
frameless ALOHA: optimizing the stopping criterion

- heuristic termination criterion:
- stop the contention when:
  - \( F_R \geq V \), or
  - \( T_I = 1 \)
- the highest reported non-asymptotic throughputs so far

<table>
<thead>
<tr>
<th>( N )</th>
<th>50</th>
<th>100</th>
<th>500</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_{GA} )</td>
<td>0.83</td>
<td>0.84</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>( T )</td>
<td>0.82</td>
<td>0.84</td>
<td>0.87</td>
<td>0.88</td>
</tr>
<tr>
<td>( F_R )</td>
<td>0.75</td>
<td>0.76</td>
<td>0.76</td>
<td>0.76</td>
</tr>
<tr>
<td>( M/N )</td>
<td>0.97</td>
<td>0.95</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>( \beta )</td>
<td>2.68</td>
<td>2.83</td>
<td>2.99</td>
<td>3.03</td>
</tr>
<tr>
<td>( V )</td>
<td>0.83</td>
<td>0.87</td>
<td>0.88</td>
<td>0.89</td>
</tr>
</tbody>
</table>
outline

- future connectivity landscape
- Ultra-Reliable Low Latency Communication
- massive Machine Type Communication (mMTC)
- at the intersection of massive and ultra-reliable
- IoT and distributed ledgers
possible when the information across devices is correlated (e.g. alarm)
the scenario

a device generates two message types
- individual update, independent of others
- alarm-type message, correlated for all sensors
model for tradeoff between massive and ultra-reliable

- the alarm set has \( M_a \) messages
- alarm occurs with probability \( p_a \)
- when alarm occurs, the message to be sent is chosen uniformly among the \( M_a \) messages

the **same alarm message** is sent by all nodes that detect it

- a node detects alarm with probability \( p_d \)
- the probability to send an alarm message is

\[
p_a p_d
\]
model for tradeoff between massive and ultra-reliable

- the standard set has $M_s$ messages
- if no alarm is detected, a node sends a standard message with probability $p_s$
- each node selects the standard message uniformly from the standard set of $M_s$ messages
- the probability to send a standard message is $p_s(1 - p_a p_d)$
information rate and spectral efficiency

there are in total $N$ devices and $K$ of them are active

example of low spectral efficiency
- low $p_s$, high $p_a$ and $p_d$ high, $N = 10$ and $K = 9$
- then most likely all nodes send the same alarm message
- the information generated in this transmission is low

example of high spectral efficiency
- high $p_s$, low $p_a$ and $p_d$ high, large $N$
error events

the probability of error in alarm decoding is much lower

error in standard message decoding occurs if two devices send the same standard message

ersors due to false positive needs to be taken into account
undetected alarm vs. spectral efficiency

Fig. 2. Trade-off between probability of error for alarm messages and the spectral efficiency. Blocklength \( n = 30000 \), \( N = 1000 \), target error probabilities \( \varepsilon_s = 10^{-1} \), \( \varepsilon_{fp} = 10^{-5} \), set sizes \( M_s = 2 	imes 10^9 \), \( M_a = 2 	imes 10^3 \), \( p_s = 0.01 \) and \( p_a = 1.0 \).

We first study the trade-off between the probability of error for alarm messages and the per-device spectral efficiency \( S \), during the event of an alarm. We consider a setting with \( N = 1000 \) devices and a blocklength of \( n = 30000 \). The alarm and standard messages are 3 and 100 bits, respectively. The probability of activation when there is no alarm is \( p_s = 0.01 \), and the transmission power is chosen such that the target average error bound for standard messages is \( \varepsilon_s = 10^{-1} \), and the probability of false positive alarms is below \( \varepsilon_{fp} = 10^{-5} \). Having only a few bits for alarm messages is a realistic setting, e.g. in a sensor network the alarm could be that a sensed value is too high or too low resulting in only one bit needed for the alarm message.

In Fig. 2 it can be seen that the probability of error increases for increasing spectral efficiency (decreasing \( p_d \)). Notice that the maximum spectral efficiency is achieved when the error probability is one (or equivalently, \( p_d = 0 \)), i.e. no alarm messages are detected. This is expected since a higher number of devices transmitting alarm messages reduces the per-device spectral efficiency, but increases the received signal-to-noise ratio of alarm messages. Furthermore, very high reliability is achievable. This trade-off between spectral efficiency and probability of error is not surprising since this is also the case when the blocklength or message set size are changed. The novelty is in the fact that it is the correlation between devices that causes the trade-off.

We now consider the minimal average transmission power, \( P_0 \), required to satisfy some target error probabilities. We assume no power restriction. That is \( p_0 = 0 \) in Theorem 3. Let all parameters be fixed except \( P_0 \) and \( p_d \). This is an optimization problem on the form:

\[
\begin{align*}
\minimize & \quad P_0, \\
\text{s.t.} & \quad P N K = 0, \\
& \quad P K (K | \neg A) (b(K) + c(K)) \leq \varepsilon_s P N K = 0, \\
& \quad P K (K | A) P K a = 0, \\
& \quad p(K a) (1 d(K, K a) (1 c(K, K a))) \leq \varepsilon_a P N K = 0.
\end{align*}
\]

The constraint functions are strictly increasing for decreasing values of \( P_0 \) and \( p_d \), and the first two constraints do not depend on...
Tradeoff EbN0 and number of devices

Fig. 3. Trade-off between $EbN0$ and the number of devices, $N$, for different values of alarm probability $p_a$ and for uncorrelated devices. Blocklength $n = 30000$, target error probabilities $e^d = f_p = 10^{-5}$, set sizes $M_s = 2^{100}$, $M_a = 2^{3}$ and $p_s = 0.01$.

**APPENDIX A**

**PROOF OF THEOREM 1**

To explicitly show the dependency on the number of messages define $T_{NK} = \{W_1, W_2, M_a[M_s] \ldots W_K, 2M_a[M_s] \ldots W_K+1, 2M_a[M_s] \ldots W_N\}$ as the event that the first $K$ out of $N$ devices transmit and the rest are silent. Due to symmetry in the devices, and without loss of generality, we assume the $K$ first devices that are transmitting. By the law of total probability this event has probability $p(T_{NK}) = p_a(p_d + (1-p_d)p_s)K (1-p_d)N-K (1-p_s)N-K + (1-p_a)p_Ks (1-p_s)N-K$.

System spectral efficiency, $S$, is defined as $S = H(W_1) / n$ where the joint entropy of all $K$ messages can be expressed using the chain rule for entropy [13, Theo. 2.5.1] as $H(W_1) = \sum_{k=1}^{K} H(W_k | W_1)$. Thus we need to express the conditional entropy $H(W_k | W_1)$ given by $H(W_k | W_1 = w_1) = \sum_{w_1} p(w_k | w_1, T_{NK}) \log_2(p(w_k | w_1, T_{NK}))$, where $H(W_k | W_1 = w_1) = \sum_{w_1} p(w_k | w_1, T_{NK}) \log_2(p(w_k | w_1, T_{NK}))$. 

IEEE Future Networks Tutorials @ 6G summit, Levi, Finland, 24-26 March 2019
summary and outlook

- the problem of massive access tightly related to the traditional problem of random access

- we have discussed several communication models, including new ones to tackle asymptotically large number of devices

- fresh breeze in the models by considering traffic correlation of IoT devices
outline

- future connectivity landscape
- Ultra-Reliable Low Latency Communication
- massive Machine Type Communication (mMTC)
- at the intersection of massive and ultra-reliable
- IoT and distributed ledgers
what is a distributed ledger?

- **ideal**: a truthful, immutable account of the history of transactions, identically understood by all participants

- **actual**: approximation of the ideal due to nonzero delay in writing in databases
what is a distributed ledger?

- **consensus** among the nodes supported by a highly nontrivial acquisition of the right to create an update.

- **smart contract**: modifying the state of the database through a predefined set of functions.

- **blockchain** is a specific instance of a distributed ledger with linear, chain structure

- The objective is to have finalized transactions, which are accepted transactions that will not be reverted with a high probability.
IoT and smart contracts

IoT connectivity

IoT smart contracts

decentralized apps

blockchain

connectivity security
blockchains and wireless IoT connectivity

- blockchains and distributed consensus fundamentally change the communication traffic
- improved security adds an additional overhead

IoT without blockchain

IoT with blockchain
what is a distributed ledger?

- **consensus** among the nodes supported by a highly nontrivial acquisition of the right to create an update.

- **smart contract**: modifying the state of the database through a predefined set of functions.

- **blockchain** is a specific instance of a distributed ledger with linear, chain structure

- the objective is to have finalized transactions, which are accepted transactions that will not be reverted with a high probability.
how DLTs fit with IoT applications

scalability

storage

application-agnostic

limited Capacity
fitting DLT to mMTC and URLLC

- capabilities and requirements of mMTC and URLLC devices quite different

- the challenge in URLLC is transaction finalization delay

- the challenge in mMTC is computational requirements and protocol overhead
IoT interactions through a blockchain

- device-to-device economic transaction
- the blockchain network acts as a third-party authority
- devices need to stay synchronized with the blockchain network.

Diagram:

3. transaction
1. ask for service
2. ask for payment
5. provide service
4. transaction confirmed

IoT device 1

IoT device 2
distributed trust architecture

Legacy System Architecture

Distributed Trust Architecture
three configurations for IoT blockchain

P1 (full node)
- The device downloads complete blocks *(header + body)* from BN

P2 (light node)
The device pre-defines a list of events of interest (= modifications of particular accounts)
It downloads block *headers* by default, but new states on demand.

P3 (delegating node)
The IoT device trusts a BN, that only forwards the events of interest.
Low security guarantees.

lightweight synchronization protocols

P2

0. Phase only in Fabric
1. digital signing
   + IoTA solves the PoW
2. forward the transaction to GateWay
3. forwarded to DLT
4. PoW to include the transaction in the ledger (not IoTA)
5., 6., 7. informing GW and IoT devices

message exchange
DLT parameters in the case study with LoRaWAN

<table>
<thead>
<tr>
<th>DLT</th>
<th>Capacity (transactions per second)</th>
<th>Average Validation Time (seconds)</th>
<th>Average Block Size (bytes)</th>
<th>Size of block header (bytes)</th>
<th>Minimum size of transaction (bytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethereum</td>
<td>1–10^2</td>
<td>1–10</td>
<td>10^3–10^4</td>
<td>508</td>
<td>110</td>
</tr>
<tr>
<td>IOTA</td>
<td>10–10^3</td>
<td>10–10^2</td>
<td>-</td>
<td>-</td>
<td>1600*</td>
</tr>
<tr>
<td>Bitcoin</td>
<td>1–10^1</td>
<td>10^2</td>
<td>10^5–10^6</td>
<td>80</td>
<td>247</td>
</tr>
<tr>
<td>Fabric</td>
<td>10^3</td>
<td>10^0</td>
<td>10^5–10^6</td>
<td>72</td>
<td>3060**</td>
</tr>
</tbody>
</table>

When using popular DLTs in LoRaWAN and IEEE 802.11 wireless networks, which are widely used for IoT.

A. Case study 1: Publishing and subscribing to a DLT via LoRaWAN

LoRaWAN is a low-power wide area network technology which enables long-range communication at modest data rates. In Europe, LoRaWAN operates in ISM bands, where it is imposed a duty-cycle of 1%, which is enforced by waiting for 99 transmission times after each transmission. This constraint limits the amount of downlink traffic from the access point (AP), challenging the suitability of LoRaWAN for devices that subscribe to information streams. We shall investigate subscribing and publishing in the following paragraphs.

We assume a 1000 meter radius cell with a single AP in the center and 100 Class C devices spread uniformly throughout. Spreading factors (SFs) are allocated based on their distance to the AP as in [14], and we account for both co- and inter-SF interference and the demodulation capabilities of the AP [14]. The UL consists of five 125 kHz UL channels in sub-band G and three in sub-band G1, within the 868 MHz ISM band with a 1% duty-cycle. UL transmissions are all of the acknowledged type with up to 3 retransmissions. In the DL the block headers are broadcasted using SF 12, that allows reaching devices at the cell edge. Under these conditions, the time to transmit a block header for Ethereum, Bitcoin and Fabric are 24.49 s, 4.11 s and 3.95 s, respectively. This corresponds to a rate limit of, respectively, 0.04, 0.24 and 0.25 block headers per second in a 100% duty-cycled band.
case study: pub/sub o a DLT via LoRaWAN

CDF and end-to-end latency

- Bitcoin
- Ethereum
- IOTA
- Hyperledger Fabric
- No DLT
case study: accountability of operations

A) authenticated gossip consensus
B) smart contract-aided computation
C) transaction-based computation
blockchain, slicing, and reliability

at present: blockchain client is treated as an application
blockchain, slicing, and reliability

proposal: reliable blockchain synchronization slice
summary and outlook

- DLTs+IoT just starting, open for protocol redesigns

- the capabilities of the edge network to be taken into account

- additional overhead and traffic pattern challenge the traditional understanding of IoT requirements