Multi-Arm Bandit Algorithms for Internet of Things Networks: A TestBed Implementation and Demonstration that Learning Helps

Lilian Besson[1,2] and Rémi Bonnefoi[1]

***Abstract*— We propose a demonstration for the ICT 2018 conference[1].**

We implement a simple network, with one Base Station, interfering traffic and several intelligent devices, communicating in a wireless protocol with slotted frequency. We model the network access as a discrete sequential decision making, and using the framework and algorithms from Multi-Armed Bandit (MAB) learning, we show that intelligent devices can improve their access to the network by using simple and decentralized algorithms such as UCB and Thompson Sampling.

I. Objectives and importance

Unlicensed bands are more and more used and considered for mobile and LAN communication standards (WiFi, LTE-U), and for Internet of Things (IoT) standards for short-range (ZigBee, Z-Wave, Bluetooth) and long-range (LoRaWAN, SIGFOX, Ingenu, Weightless) communications [1]. Efficient Medium Access (MAC) policies allow devices to avoid interfering traffic and can significantly reduce the spectrum contention problem in unlicensed bands. As end-devices battery life is a key constraint of IoT networks, this leads to IoT protocols using as low signaling overhead as possible and simple ALOHA-based mechanisms. In this demo, we evaluate Multi-Armed Bandits algorithms [2], used in combination with a pure ALOHA-based protocol. We consider the Upper-Confidence Bound (UCB) [3], and the Thompson-Sampling (TS) algorithms [4]. Both algorithms have already been applied with success, for Opportunistic Spectrum Access [5] and recently for multi-users Cognitive Radio problems [6].

The aim of this demo is to assess the potential gain of learning algorithms in IoT scenarios. In a very simple wireless network, consisting of one Base Station Gateway (BTS) and a certain interfering background traffic assumed to be stationary, some dynamic intelligent devices (end-user, EU) try to access the network, with a simple low-overhead protocol. To simulate networks designed for the Internet of Things (IoT), we consider a protocol with no sensing, no repetition of uplink messages, and where the BTS is in charge of sending back an acknowledgement, after some fixed-time delay, to any EU who succeeded in sending successfully an uplink packet. By considering a small number of wireless channels and only one PHY layer configuration (*i.e.*, modulation, waveform, etc), and in case of a non-uniform traffic in the different channels, the EU can improve their usage of the network if they are able to *learn* on the fly the best channels to use (*i.e.*, the most vacant).

Following our recent work [7], we propose to model this problem as Non-Stationary Multi-Armed Bandit, and suggest to use low-cost algorithms, focusing on two well-known algorithms: a frequentist one (UCB, Upper Confidence Bounds) and a Bayesian one (Thompson Sampling). We use a TestBed designed in 2017 at SCEE [8], containing different USRP cards [9], controlled by a single laptop using GNU Radio [10], and where the intelligence of each EU corresponds to a learning algorithm, implemented as a GNU Radio Companion block [11] and written in Python or C++.

Our demo can modify dynamically the background traffic, and reset at any time the channel selection algorithms of 4 End-Users with low duty-cycle, one running the naive uniform access and the rest running different learning algorithms. This allow to check that in case of uniform traffic, there is nothing to learn and the intelligent EU do not reduce their successful communication rate in comparison to the naive EU, and in case of stationary non-uniform traffic, the MAB learning algorithms indeed help to increase the global efficiency of the network by improving the success rate of the 3 intelligent EU.

*Both authors contributed equally and are ordered alphabetically.

[1]CentraleSupélec (campus of Rennes), IETR, SCEE Team, Avenue de la Boulaie - CS 47601, 35576 Cesson-Sévigné, France, France Email: lilian.besson, remi.bonnefoi at centralesupelec.fr

[2]Univ. Lille 1, CNRS, Inria, SequeL Team, UMR 9189 - CRIStAL, F-59000 Lille, France Email: lilian.besson at inria.fr

[1]In reply of ict-2018.org/call-for-demos/

II. Timeline

The design of our demo is fixed, and so far everything works, except we need to finish debugging the synchronisation between the different dynamic devices and the base station. We plan to finish this by January 2018, and if we have time after, we will also consider to add another base station in the system.

III. Demo layout

Figure 1 shows the layout of our demonstration. Figures 2, 3 and 4 show the GNU Radio Companion (GRC, [11]) schemes corresponding respectively to the Interfering Traffic generator (in charge of generating a random stationary traffic in each channel, with a fixed duty-cycle), the Base Station (in charge of listening in each channel, detecting incoming messages and replying with an acknowledgement), and the Dynamic Device (in charge of emitting at a low duty-cycle in a sequentially chosen channel, receiving an acknowledgement and using the statistics and its learning algorithm to decide which channel to use next). In our demo, the dynamic devices try to communicate with the base station. This communication is hindered by some interfering traffic. This interfering traffic is supposed to be unevenly distributed in channels and is generated by one USRP. If the base station receives a message transmitted by an intelligent end-device, it acknowledges the packet.

The basic blocks (sink, source, FFT) are builtin GRC, but all the other blocks (`Demodulator`, `send_ack`, `Check_ack` for the BTS, `Generator` for the interfering traffic, and `Renormalize_ack` and `generator_SU` for the intelligent devices) are written in Python or C++ for this demonstration.

IV. List of equipment for demo

We require a large-screen TV, but we will bring everything else. Our demo is developed using a large Testbed, as showed in Figure 5, but it can be transported by using only one laptop, connected to a portable switch, to control the 6 different USRP cards.

Our demo operates on the $433.5\,\mathrm{MHz}$ band, with a bandwidth of $2\,\mathrm{MHz}$.

V. Conclusion

Possible extensions of this work include: considering any number of dynamic device and not only 4, implementing a real-world IoT communication protocol (like the LoRaWAN standard), and studying the interference in case of a second Base Station located nearby.

Acknowledgment

The authors would like to thank Christophe Moy for his insights, and we acknowledge the work of two CentraleSupélec students, Clément Barras and Théo Vanneuville, as we took inspiration in the GNU Radio code they wrote during their project in Spring 2017.

This work is supported by the French National Research Agency (ANR), under the projects SOGREEN (grant coded: *N ANR-14-CE28-0025-02*) and BADASS (*N ANR-16-CE40-0002*), by Région Bretagne, France, by the CNRS, under the PEPS project BIO, by the French Ministry of Higher Education and Research (MENESR) and ENS Paris-Saclay.

References

[1] M. Centenaro, L. Vangelista, A. Zanella, and M. Zorzi, "Long-range communications in unlicensed bands: the rising stars in the IoT and smart city scenarios," *IEEE Wireless Communications*, vol. 23, no. 5, pp. 60–67, 2016.

[2] S. Bubeck, N. Cesa-Bianchi, *et al.*, "Regret analysis of Stochastic and Non-Stochastic Multi-Armed Bandit Problems," *Foundations and Trends® in Machine Learning*, vol. 5, no. 1, pp. 1–122, 2012.

[3] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time Analysis of the Multi-armed Bandit Problem," *Machine Learning*, vol. 47, no. 2, pp. 235–256, 2002.

[4] W. R. Thompson, "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, vol. 25, 1933.

[5] W. Jouini, D. Ernst, C. Moy, and J. Palicot, "Upper Confidence Bound Based Decision Making Strategies and Dynamic Spectrum Access," in *2010 IEEE International Conference on Communications*, pp. 1–5, 2010.

[6] V. Toldov, L. Clavier, V. Loscr, and N. Mitton, "A Thompson Sampling approach to channel exploration-exploitation problem in multihop cognitive radio networks," in *PIMRC*, pp. 1–6, 2016.

[7] R. Bonnefoi, L. Besson, C. Moy, E. Kaufmann, and J. Palicot, "Multi-Armed Bandit Learning in IoT Networks: Learning helps even in non-stationary settings," in *12th EAI Conference on Cognitive Radio Oriented Wireless Network and Communication*, CROWNCOM Proceedings, 2017.

[8] Q. Bodinier, *Coexistence of Communication Systems Based on Enhanced Multi-Carrier Waveforms with Legacy OFDM Networks*. PhD thesis, CentraleSupélec, 2017.

[9] "USRP Hardware Driver and USRP Manual." `http://files.ettus.com/manual/page_usrp2.html`. Accessed: 2017-12-15.

[10] "GNU Radio Documentation." `https://www.gnuradio.org/about/`. Accessed: 2017-12-15.

[11] "GNU Radio Companion Documentation." `https://wiki.gnuradio.org/index.php/GNURadioCompanion`. Accessed: 2017-12-15.

Note: the code of our demo is for GNU Radio [10] and GNU Radio Companion [11], and is open-sourced under the GPLv3 License, on Bitbucket at: `https://goo.gl/U7d8vS`.

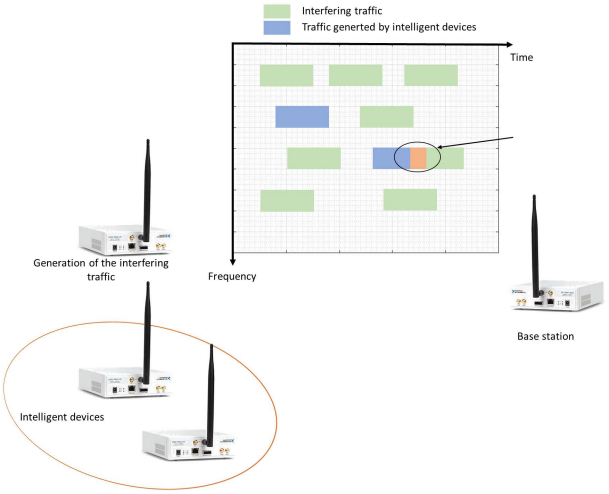

Fig. 1: Layout of the demonstration. Our demo uses one USRP to generate an interfering traffic and one for the base station. Moreover, at least one intelligent device is added in the network and tries to communicate with the base station. The protocol is a pure ALOHA protocol with acknowledgement.

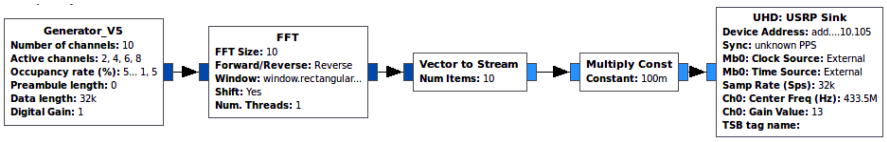

Fig. 2: GRC Scheme for the Interfering Traffic (TX PU): 1 USRP block (sink).

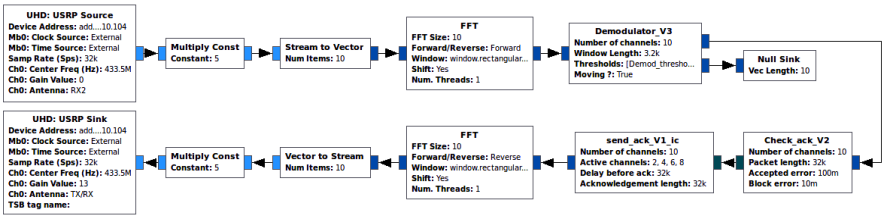

Fig. 3: Base Station (RX/TX BTS): 2 USRP blocks (sink/source) but one card.

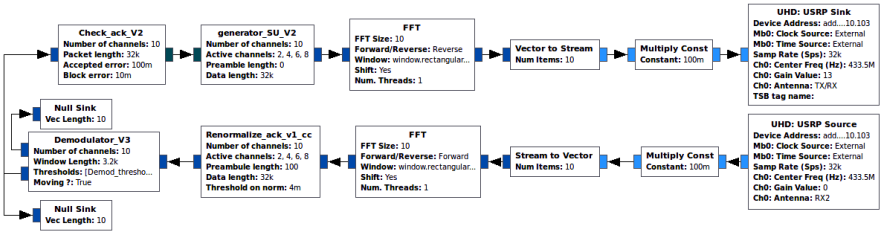

Fig. 4: Dynamic Device (RX/TX SU): 2 USRP blocks (sink/source) but one card.

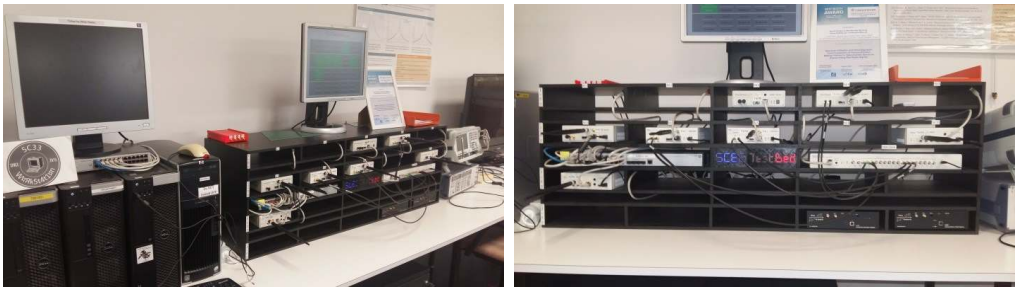

Fig. 5: Our TestBed, cf. `http://www-scee.rennes.supelec.fr/wp/testbed/`.