Multi-Armed bandit Learning in Iot Networks (MALIN)

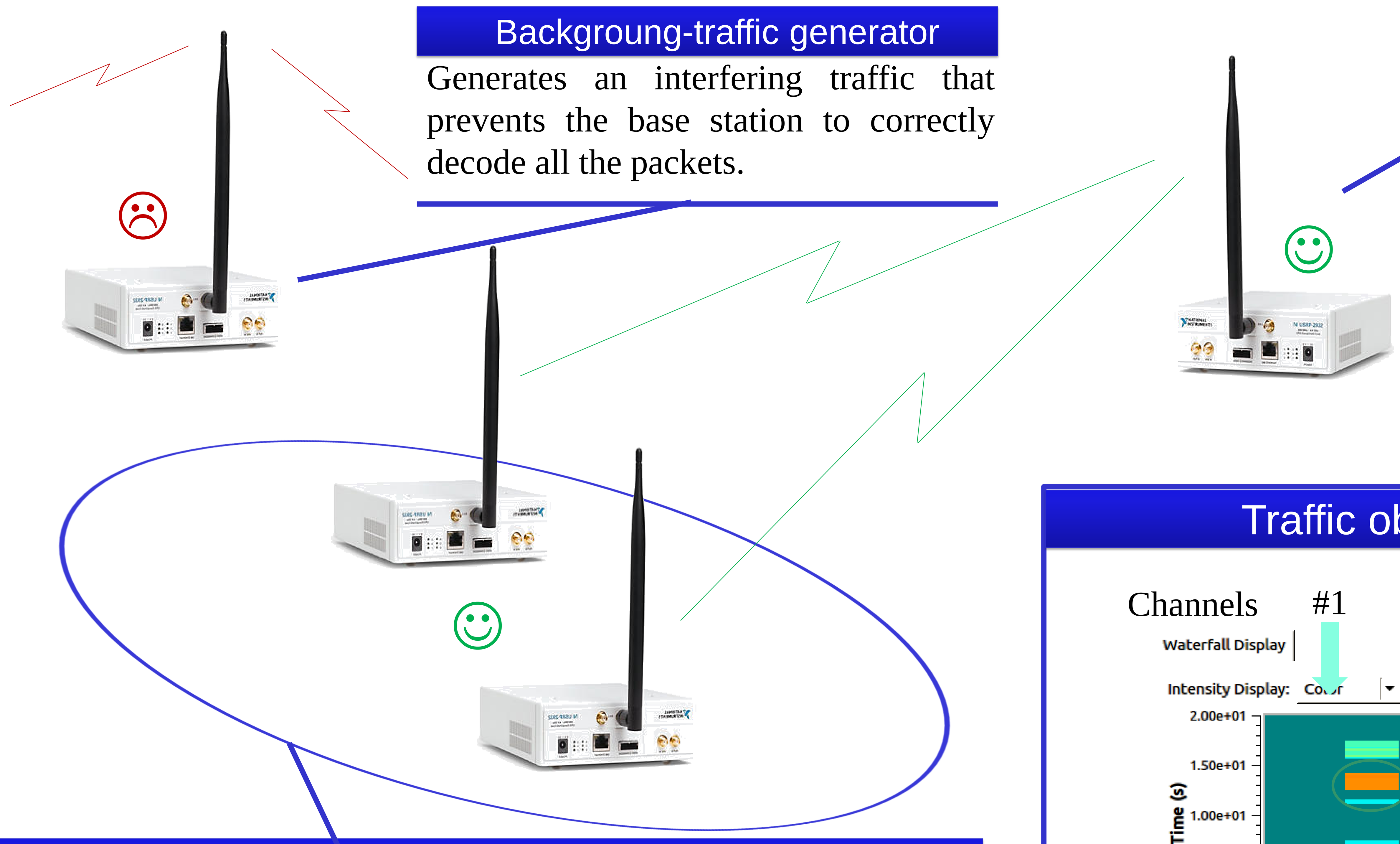
Rémi Bonnefoi[1], Lilian Besson[1] and Christophe Moy[2]

[1] CentraleSupélec/IETR, F-35576, Cesson-Sévigné Cedex, France. *Remi.Bonnefoi@CentraleSupelec.fr, Lilian.Besson@CentraleSupelec.fr*

[2] Univ Rennes, CNRS, IETR – UMR 6164, F-35000, Rennes, France. *Christophe.Moy@Univ-Rennes1.fr*

Goal

- With the advent of the Internet of Things (IoT), unlicensed band are going to be shared by a large number of devices with dissimilar caracteristics. In such context, solutions are required to allow the coexistence of devices and to avoid performance drop due to interference.
- In this demonstration, we show that reinforcement learning algorithms and in particular Multi-Armed Bandit algorithms can be used as a means of improving the performance of IoT communications.



Backgroung-traffic generator

Generates an interfering traffic that prevents the base station to correctly decode all the packets.

Gateway

Receives and decodes the packets sent by our intelligent devices in different channels.

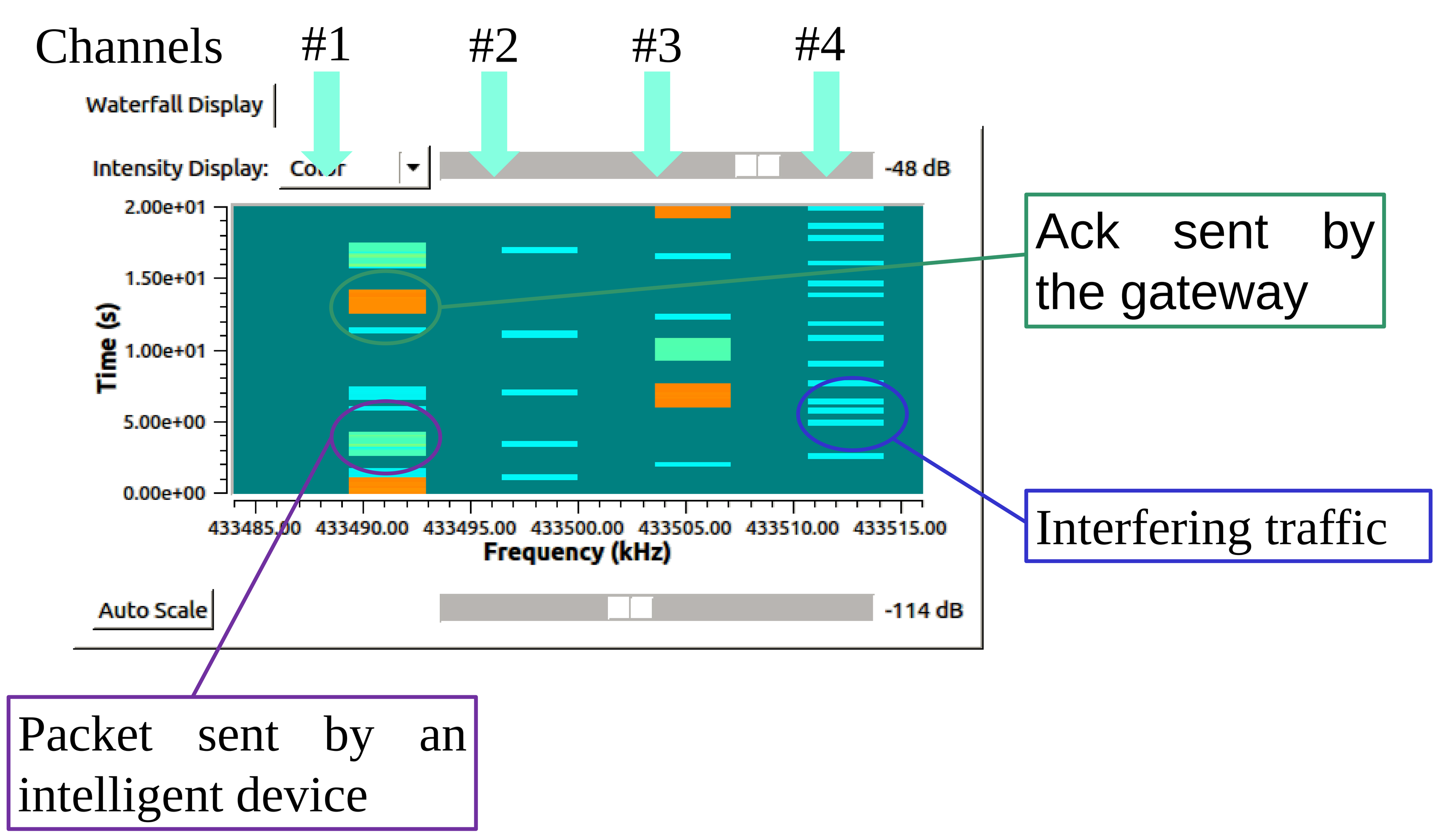
Once a packet is successfully received, the gateway sends an acknowledgement in the channel used for the uplink packet

Intelligent devices

They want to send packets to the gateway. For that purpose, they can use the different channels available. The selection of the channel is done using a MAB learning algorithm in order to avoid collisions with other objects transmissions.

Each objet makes its own decision. They do not share any information. The decision is decentralized and uncoordinated.

Traffic oberved by the gateway



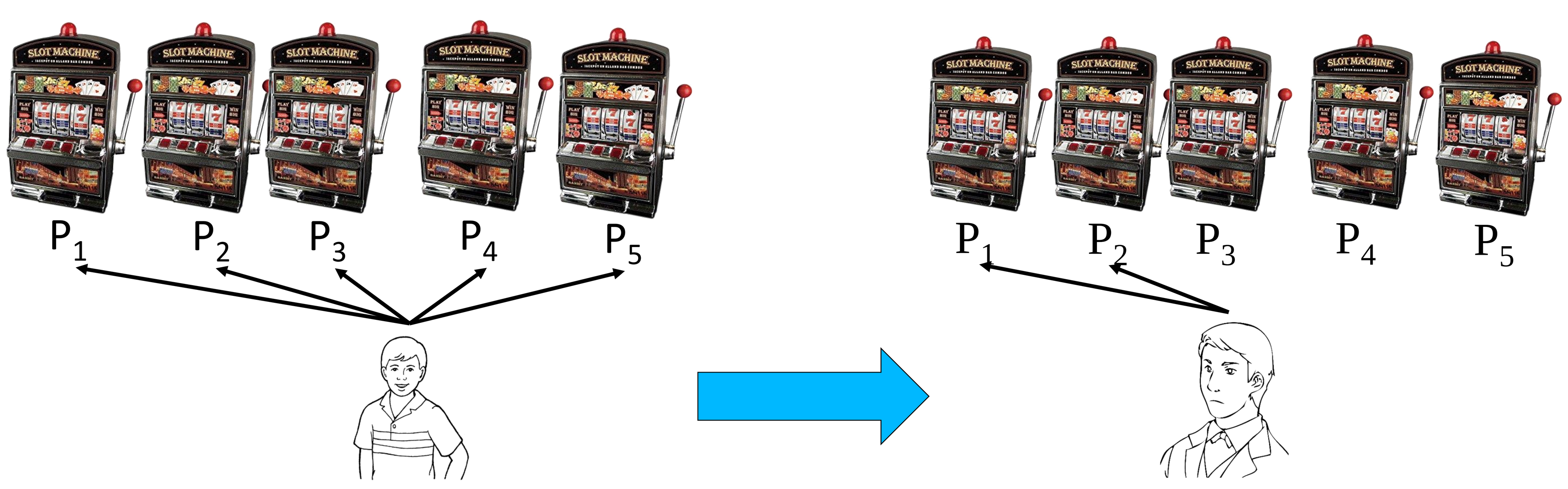
MAB learning

- A user faces N choices (*e.g.* N channels)
- The channels provide him a reward with a given probability

How to identify the best channel?

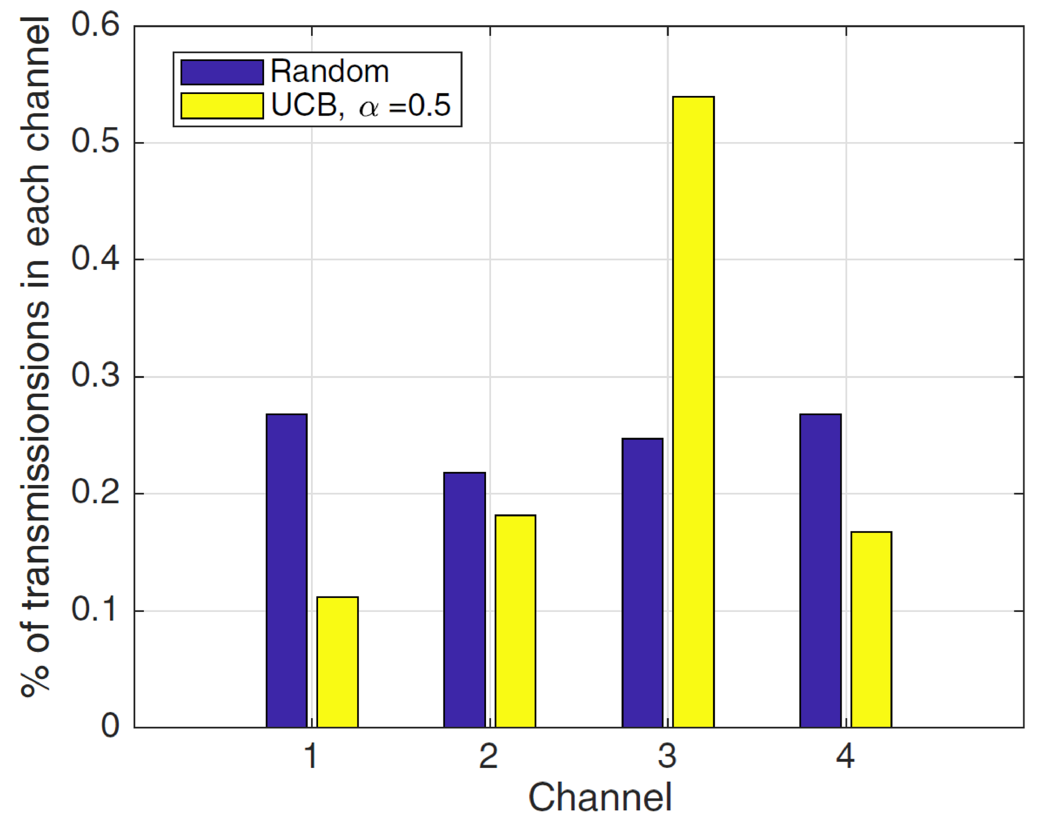
- MAB learning algorithms are proved to be optimal to solve this problem.
- Any MAB algorithm can be used for channel selection (UCB, Thompson Sampling or others)
- With the UCB algorithm, the channel with the highest index is chosen for each transmission

$$B_j(t) = \frac{\sum_{l=0}^{t-1} r_l(t)\mathbb{1}(a_l = j)}{T_j(t)} + \sqrt{\frac{\alpha \ln(t)}{T_j(t)}}$$



Results

Channels occupancy rate [20%, 10%, 5%, 25%]



Acknowledgement

This work is supported by the European Union through the European Regional Development Fund (ERDF), and by Ministry of Higher Education and Research, Brittany and Rennes Métropole, through the CPER Project SOPHIE / STIC & Ondes. Part of this work is also funded by the ANR projects SOGREEN and BADASS, by Région Bretagne, France, by the French Ministry of Higher Education and ENS Paris Saclay.

[1] R. Bonnefoi, L. Besson, C. Moy, E. Kaufmann, J. Palicot, Multi-Armed Bandit Learning in IoT Networks: Learning Helps Even in Non-stationary Settings. In CrownCom 2017.

[2] S. Bubeck and N. Cesa-Bianchi, “Regret analysis of stochastic and nonstochastic multiarmed bandit problems,” *Foundations and Trends® in Machine Learning*, vol. 5, no. 1, pp. 1–122, 2012.