Titre : Secure AIoT

Supervisor :

Van-Tam Nguyen (van-tam.nguyen@telecom-paris.fr)

Summary:

AIoT describes the convergence of IoT and AI systems with the common goal of generating useful IoT data and processing that data as early as possible. The main objective of this master II internship is to overcome the limitations of computing power and memory at the node level in a decentralized IoT network, by embedding them with Artificial Intelligence (AI) integrated with game-theoretic decision analytics.

Description:

AI is essential to the success of the IoT. Traditional methods can no longer be applied to IoT deployments where there are huge volumes of IoT data being generated at a very high rate. So far, AI has mainly been found in two places within the IoT system: the center and the “edge”. AI near IoT nodes will improve security, privacy, and help reduce latency and bandwidth.

AIoT describes the convergence of IoT and AI systems with the common goal of generating useful IoT data and processing that data as early as possible. The impact of the combination of IoT and AI can be huge. AIoT not only brings the compute closer to where the data resides, but adds the intelligence needed to improve reliability, efficiency and productivity.

Current IoT architecture emphasizes centralized information processing, with the connected objects mostly serve as data collection nodes. As the number of connected objects increases exponentially, the centralized architecture of IoT networks suffers from major bottlenecks, including connectivity (interference), in latency (distant servers), in energy consumption (cost of distant communication), in centralized data processing (data flooding) and in centralized network management (number of objects). To address these bottlenecks, in this internship, we will propose two major changes the current IoT architecture. First, we will propose to make network management and computation decentralized, by enabling the connected objects (the nodes) become active elements and perform distributed tasks within the network including network sensing, connectivity decision, security verification, data compression, data pruning and distributed calculation. Second, we will empower the intelligence of the individual nodes by embedding them with Artificial Intelligence (AI) integrated with game-theoretic decision analytics. AI embedded into IoT nodes will improve security, privacy, and help reduce latency and bandwidth. AIoT not only brings the compute closer to where the data are generated, but adds the intelligence needed to improve robustness, reliability, efficiency, availability and productivity.

The major challenge to the proposed architecture is the limitations of the computing resources at the nodes. Most IoT networks rely on low-cost, low-energy IoT devices, which have a limited memory, computing power and storage . The on-chip memory may be 3 orders of magnitude smaller than mobile devices, and 5-6 orders of magnitude smaller than cloud GPUs, making top-performance learning algorithms (such as deep learning) deployment impossible. In addition, they may not have DRAM nor the OS and the tight memory budget will limit the input size. Finally, these embedded IoT devices are

opening opportunities for intruders to commit cyber-attacks i.e., black hole, gray hole, DDoS attacks. Therefore, new methods are required to counter them in future devices.
To overcome the limitations of computing power and memory at the node level in a decentralized IoT network, we propose the following research tasks:

1) Development of deep learning algorithms using attention-based architecture (such as Transformer) for resource-constrained systems. Attention-based architecture can be designed to focus on important information and can capture temporal sequences, which then can be integrated with game-theoretic decision analytics for driving decisions at the node level, including security decisions, or relying on Knowledge Graph embedding;
2) Development of algorithms for model compression, including pruning, quantization, coding, approximative computing;
3) Design of energy-saving algorithms taking into account the energy cost of data movements;
4) Inference on patches/blocks, co-design of architecture and planning of inferences;
5) Memory distribution analysis and optimization, hardware and algorithmic co-design.