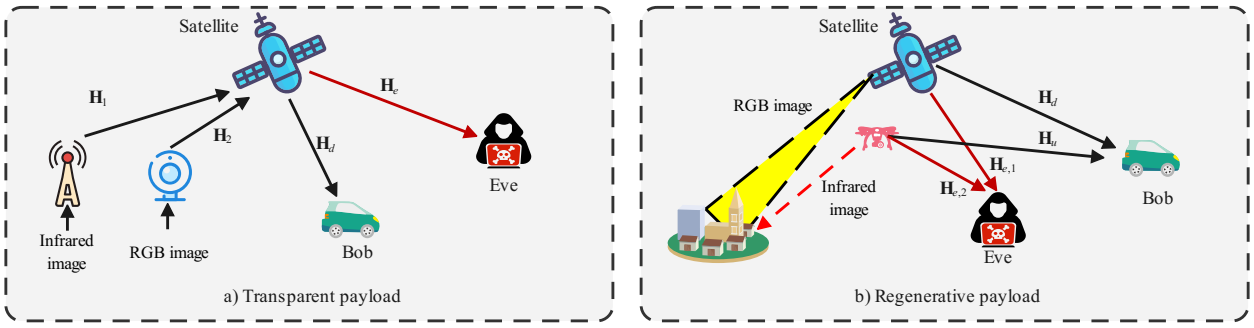

Fig. 1: NTN system model with the transparent payload and regenerative payload modes.

In the paper, we investigate an NTN channel model, illustrated in Fig. 1. In the transparent payload, the model comprises a satellite equipped with N_t antennas, an Eve equipped with N_e antennas, an N_a-antenna infrared sensor, an N_a-antenna RGB sensor, and a legitimate user (Bob) with N_r antennas. Assuming Rayleigh fading for wiretap channels, the downlink channel between the satellite and Eve is represented as $\mathbf{H}_e \sim \mathcal{CN}_{N_e,N_t}(\mathbf{0}, \mathbf{I}_{N_e} \otimes \mathbf{I}_{N_t})$. Meanwhile, the channels from the satellite to Bob, from the infrared sensor to the satellite, and from the RGB sensor to the satellite are denoted as $\mathbf{H}_d \in \mathbb{C}^{N_r \times N_t}$, $\mathbf{H}_1 \in \mathbb{C}^{N_t \times N_a}$ and $\mathbf{H}_2 \in \mathbb{C}^{N_t \times N_a}$, respectively. For simplicity, it is assumed that $\mathbf{H}_d$, $\mathbf{H}_1$, and $\mathbf{H}_2$ are perfectly estimated.

In the transparent payload mode, the satellite works solely as a relay without additional processing, and the semantic information is generated by two sensors, so the received signal at Bob is

$$\mathbf{y}_t = \mathbf{H}_d(\mathbf{H}_1\mathbf{F}_1\mathbf{x}_1 + \mathbf{H}_2\mathbf{F}_2\mathbf{x}_2 + \mathbf{n}_s) + \mathbf{n}_t, \tag{1}$$

where $\mathbf{x}_1 \in \mathbb{C}^{N_a \times 1}$ and $\mathbf{x}_2 \in \mathbb{C}^{N_a \times 1}$ are the semantic information through encoding-decoding neural network with inputs $\mathbf{m}_1$ and $\mathbf{m}_2$. $\mathbf{m}_1$ and $\mathbf{m}_2$ are the infrared and RGB images, respectively. $\mathbf{F}_1 \in \mathbb{C}^{N_a \times N_a}$ and $\mathbf{F}_2 \in \mathbb{C}^{N_a \times N_a}$ are the precoding matrices outputted by the multi-modal fusion neural network. The transmit power constraints include $\mathbb{E}(|\mathbf{F}_1\mathbf{x}_1|^2) = \rho_s\mathbf{I}_{N_a}$ and $\mathbb{E}(|\mathbf{F}_2\mathbf{x}_2|^2) = \rho_s\mathbf{I}_{N_a}$. $\mathbf{n}_t$ and $\mathbf{n}_s$ correspond to the additive white Gaussian noise (AWGN) at Bob and the satellite, respectively, which obey $\mathcal{CN}_{N_r,1}(\mathbf{0}, \sigma_t^2\mathbf{I}_{N_r})$ and $\mathcal{CN}_{N_t,1}(\mathbf{0}, \sigma_s^2\mathbf{I}_{N_t})$, respectively.

In the regenerative payload, the difference is that the source of the infrared image is the UAV equipped with N_u antennas, and the source of RGB image remains the satellite. Eve can intercept messages from the UAV and satellite. The wiretap channels from the UAV to Eve and from the satellite to Eve are Rayleigh channel, which is denoted by $\mathbf{H}_{e,1} \sim \mathcal{CN}_{N_e,N_u}(\mathbf{0}, \mathbf{I}_{N_e} \otimes \mathbf{I}_{N_u})$ and

$\mathbf{H}_{e,2} \sim \mathcal{CN}_{N_e,N_t}(\mathbf{0}, \mathbf{I}_{N_e} \otimes \mathbf{I}_{N_t})$, respectively. The semantic information is generated by both satellite and UAV, so the received signal at Bob is

$$\mathbf{y}_r = \mathbf{H}_d \mathbf{F}_r \mathbf{x}_r + \mathbf{H}_u \mathbf{F}_u \mathbf{x}_u + \mathbf{n}_r, \tag{2}$$

where $\mathbf{x}_r \in \mathbb{C}^{N_t \times 1}$ and $\mathbf{x}_u \in \mathbb{C}^{N_u \times 1}$ are the semantic information through encoding-decoding neural network with inputs $\mathbf{m}_1$ and $\mathbf{m}_2$. $\mathbf{m}_1$ and $\mathbf{m}_2$ are the infrared and RGB images, respectively. $\mathbf{F}_r \in \mathbb{C}^{N_t \times N_t}$ and $\mathbf{F}_u \in \mathbb{C}^{N_u \times N_u}$ are the precoding matrices outputted by the multimodal fusion neural network. The transmit power constraints include $\mathbb{E}(|\mathbf{F}_r \mathbf{x}_r|^2) = \rho_s \mathbf{I}_{N_t}$ and $\mathbb{E}(|\mathbf{F}_u \mathbf{x}_u|^2) = \rho_u \mathbf{I}_{N_u}$.

The encoding-decoding neural network and multimodal fusion neural network are jointly trained same as in reference [5]. With trained neural networks, the sensors broadcast the outputs of neural networks as semantic signals, i.e., $\mathbf{F}_1 \mathbf{x}_1$ and $\mathbf{F}_2 \mathbf{x}_2$, to satellite, and satellite relays the signals to Bob in the transparent payload mode. In the regenerative payload mode, the satellite and UAV broadcast semantic signals $\mathbf{F}_r \mathbf{x}_r$ and $\mathbf{F}_u \mathbf{x}_u$ to Bob, respectively. We employ artificial noise methodologies to ensure the security of the neural network parameters.