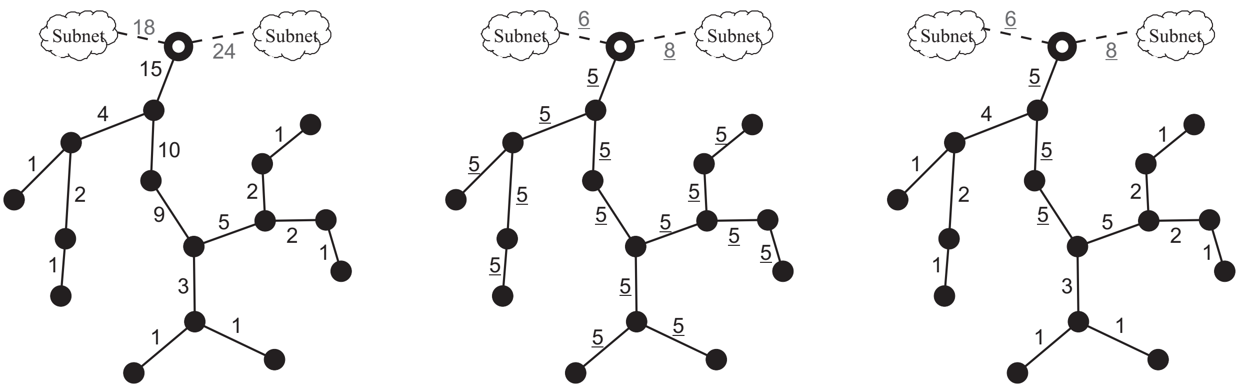

(a) Non-CS (b) Plain-CS (c) Hybrid-CS

Fig. 1. Comparison of different data collection mechanisms. The link labels correspond to the carried traffic. The underlined labels indicate CS encoded traffic, assuming $\rho = 3$. The sink is represented by the hollow circle.

the δ children and having the same number of nodes $\tilde{n} = \frac{n}{\delta}$ (of course building such a partition may not always be possible). To apply CS on each subtree we need $\tilde{n} \geq n_{\min}$. In that case, each subtree is responsible for sending $\tilde{k} = \frac{\tilde{n}}{\rho} < \frac{n}{\rho}$ packets and the sink will then receive in total k packets as before, but intermediate nodes in the network will take a much lower load ($\tilde{k} = \frac{k}{\delta}$ instead of k). We are not claiming that the best solution is necessarily to create a balanced partition of δ subnets even though we suspect that probably it is often true.

In the problem formulation described later, we will partition the network into disjoint subnets and let individual subnets aggregate data samples independently of the other subnets. Such a partition is valid as far as the size of each subnet is not smaller than $n_{\min}$. We illustrate an unbalanced partition (in particular on one of the subtrees) in Fig. 1(c). Note that whereas the CS operation (along with routing) is done independently on each subnet, the link scheduling should still be performed globally, as the interference generated by a link (no matter which subtree it belongs to) has a global impact on the rest of the network. This makes our optimization problems hard to solve even if we assume that the routing is determined by predefined subtrees.

III. System Model and Problem Formulation

In this section, we first define the models for various components of a WSN, then we present the problem formulations.

A. Models and Assumptions

We model the WSN as a set $\mathcal{N}$ of nodes, with $|\mathcal{N}| = n$, and a sink Θ. Each node $i \in \mathcal{N}$ is associated with a geographical location. We assume that (i) all nodes send sensory data (through multihop routing if necessary) to the sink with the same rate λ, (ii) time is slotted and all the nodes are synchronized, and (iii) the network is operated in a conflict-free and scheduled manner.

We assume that all the nodes have the same *transmit power* P_{tx} and the same *data-rate* c. This is only for ease of exposition; our approach does accommodate multiple powers and rates. We assume that the channel gain from a node i to another node j is quasi-static, since we consider fixed wireless networks. For simplicity, we model the channel gain as isotropic path-loss given by $(\frac{d_{ij}}{d_0})^{-\eta}$ where d_{ij} denotes the *distance* from node i to node j, d_0 is the *near-field crossover distance* and η is the *path-loss exponent*. The feasibility of a wireless link is based on whether a bit-error-rate (BER) less than a tolerable maximum can be achieved on the link. We assume that this BER requirement translates into a minimum SINR (signal-to-interference-and-noise ratio) requirement corresponding to an SINR threshold β. We define $\mathcal{L}$ as the set of all feasible links. Specifically, a link $l = (i, j)$ is feasible (or $l \in \mathcal{L}$) if $\frac{P_{\text{tx}}}{N_0}(\frac{d_{ij}}{d_0})^{-\eta} \geq \beta$ where N_0 is the *thermal noise power* in the frequency band of operation. Let $|\mathcal{L}| = L$, and let l_O and l_D denote the origin and destination of link l, respectively.

We use the following SINR-based interference model. Let $\boldsymbol{\zeta} \subset \mathcal{L}$ denote a set of links. When all the links in $\boldsymbol{\zeta}$ are simultaneously active, the SINR perceived by link $l \in \boldsymbol{\zeta}$ is given by

$$\gamma_l(\boldsymbol{\zeta}) = \frac{P_{\text{tx}}(\frac{d_{l_O l_D}}{d_0})^{-\eta}}{N_0 + \sum_{k \in \boldsymbol{\zeta} \setminus \{l\}} P_{\text{tx}}(\frac{d_{k_O l_D}}{d_0})^{-\eta}} \tag{3}$$

We say a set of links $\boldsymbol{\zeta}$ is an *independent set* (ISet) if no two links share the same node and, for every link $l \in \boldsymbol{\zeta}$, we have $\gamma_l(\boldsymbol{\zeta}) \geq \beta$. It is clear that all the links belonging to an ISet can be scheduled at the same time in a conflict-free fashion. We define $\mathcal{I}$ to be the collection of all ISets

$$\mathcal{I} = \{\boldsymbol{\zeta} | \gamma_l(\boldsymbol{\zeta}) \geq \beta, \ \forall\, l \in \boldsymbol{\zeta}\} \tag{4}$$

Let $\mathcal{I}_l$ denote the set of ISets that contain link l. We use the SINR-based interference model rather than other more frequently used ones (e.g., protocol model) simply because it is more realistic [12].

Let $\mathcal{S}$ denote the power set of $\mathcal{L}$. A *transmission schedule* is an $|\mathcal{S}|$-dimensional vector $\hat{\alpha} = [\alpha_{\boldsymbol{\zeta}}]_{\boldsymbol{\zeta} \in \mathcal{S}}$, and we can interpret $\alpha_{\boldsymbol{\zeta}}$ as the fraction of time allocated to a link set $\boldsymbol{\zeta}$. To make a schedule conflict-free, we need $\alpha_{\boldsymbol{\zeta}} > 0$ only if the set $\boldsymbol{\zeta}$ is an ISet (otherwise $\alpha_{\boldsymbol{\zeta}} = 0$) and $\sum_{\boldsymbol{\zeta} \in \mathcal{I}} \alpha_{\boldsymbol{\zeta}} \leq 1$. Therefore,