Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Resource and Network Management for Satellite Communications Systems: A Chance-Constrained Approach

Yuma Abe [*,**] **Masaki Ogura** [***] **Hiroyuki Tsuji** [*]
Amane Miura [*] **Shuichi Adachi** [**]

[*] *National Institute of Information and Communications Technology, Tokyo, Japan (e-mail: yuma.abe@nict.go.jp, tsuji@nict.go.jp, amane@nict.go.jp)*
[**] *Keio University, Yokohama, Japan, (e-mail: yuma.abe@keio.jp, adachi@appi.keio.ac.jp)*
[***] *Osaka University, Osaka, Japan. (e-mail: m-ogura@ist.osaka-u.ac.jp)*

Abstract: In this paper, we propose an optimization framework for the resource and network management in large-scale satellite communications (SATCOM) systems with stochastic and time-varying communication requests. Because further expansions of SATCOM systems are expected in the near future, it is of practical importance to develop management methodologies for efficiently and robustly operating large-scale SATCOM systems. To address this problem, in this paper, we formulate a SATCOM network optimization problem within the framework of the chance-constrained model predictive control. In this problem, we require that a joint chance constraint on the bandwidth loss rate is satisfied with a user-specified performance and probability. Because the joint chance constraint is nonlinear in design variables, we relax the constraint to a linear deterministic one for reformulating the original problem into a relaxed deterministic problem. We numerically verify that the proposed method outperforms a baseline methodology and therefore allows us to efficiently manage large-scale SATCOM systems.

Keywords: Satellite communications, Communication control applications, Communication networks, Resource and network management, Predictive control, Stochastic programming

1. INTRODUCTION

We have seen a massive expansion of satellite communications (SATCOM) systems over the past decades (del Portillo et al., 2019). This is because SATCOM systems have been required to provide increasingly larger capacity of communication and to meet various novel types of communication needs arising from the Internet of Things (IoT) (Akyildiz and Kak, 2019). Therefore, several companies have begun constructing SATCOM systems which consist of hundreds of thousands of communication satellites (Liu et al., 2018). Equipped with such a massive number of satellites, large-scale SATCOM systems can cover land, sea, and air globally and, therefore, provide broadband communication links for a large number of users. However, there is a lack of effective methodologies for the efficient management of large-scale SATCOM systems.

Two major difficulties are present in the management of large-scale SATCOM systems. The first one is the heterogeneity of components in SATCOM systems. For example, large-scale SATCOM systems contain satellites having different sizes (e.g., large, small, or nano), orbits (e.g., geostationary or non-geostationary earth orbit (GEO or NGEO)), and frequency (e.g., Ku-band, Ka-band, or optical) (Liu et al., 2018). Furthermore, the spatially distributed users may have qualitatively different communication requests depending on their specific communication needs such as mobile, emergency, and IoT communications. Therefore, the operator of the SATCOM systems has to construct a network with a large number of different types of system components.

The second difficulty is the time-variability and stochasticity of communication requests from users. In general, communication requests arising in SATCOM systems vary in time (Vasavada et al., 2016). A part of requests changes deterministically and can be estimated; for example, the requests of mobile communications from aircrafts can be roughly estimated by using their trajectories. However, the other part of the requests changes stochastically due to other unknown reasons and behaves like a non-stationary process (Bacco et al., 2018; Donner et al., 2010). Therefore, it is not trivial to estimate the stochastic part of the future behavior of the change. For these reasons, the SATCOM operator should allocate communication bandwidth resources to each user by combining a predictive approach to the deterministic change and a robust approach to the stochastic change.

In this paper, we study a resource and network management methodology for the large-scale SATCOM systems to address the aforementioned two difficulties. Stochastic and time-varying requests of users are modeled by an auto-regressive integrated moving-average (ARIMA) model, which enables us to represent a wide class of non-stationary stochastic processes (Box et al., 2015). We specifically formulate a network optimization problem using a chance-constrained model predictive control (CC-MPC) approach (Nemirovski and Shapiro, 2006; Maciejowski, 2000) to design a time-varying SATCOM network capable of dealing with communication requests stochastically arriving from users. Because the nonlinearity of the chance constraint prevents us from efficiently solving the network optimization problem, we perform a relaxation of the constraint to a linear deterministic one which in turn allows us to efficiently solve the network optimization problem sub-optimally. As a result of solving the optimization problem, the SATCOM operator obtains a desired set of resource allocation and network structure that is robust to the stochasticity of user requests. We finally demonstrate the effectiveness of the proposed method by numerical simulations.

Notation: The symbol k denotes the discrete time. The symbols $\mathbb{R}, \mathbb{R}^n, \mathbb{R}^{n\times m}$ and $\{0,1\}^{n\times m}$ denote sets of real numbers, n-dimensional real vectors, $n \times m$ real matrices, and $n \times m$ binary matrices, respectively.

2. SYSTEM MODEL

In this section, we describe our model of the large-scale SATCOM systems.

2.1 Large-Scale SATCOM Systems

Figure 1 shows the large-scale SATCOM systems that we study in this paper. User terminals (users) such as aircraft and ships request communication links to communication satellites; and the satellites provide the links for the users. Communication data from the users are aggregated in the satellites and sent to gateway stations (gateways), which are connected to the ground network. If some satellites cannot directly send communication data to the gateways due to bandwidth limitations, the satellites are allowed to communicate with another satellite to reroute the data. Communication links between user-satellite, satellite-satellite, and satellite-gateway are called user links, inter-satellite links, and feeder links, respectively. In this paper, we assume that the user links and the feeder links are bidirectional and the inter-satellite links are directional. A SATCOM operator is expected to design a network of user links, inter-satellite links, and feeder links flexibly to meet the users communication requests by managing the large number of the satellites and gateways. We collectively call the user link, inter-satellite link, and feeder link network as a SATCOM network.

Let $N_{\mathrm{U}}, N_{\mathrm{S}}$, and N_{G} be the number of users, satellites, and gateways in the large-scale SATCOM systems, respectively. We denote index sets of the users, satellites, and gateways as $\mathcal{N}_{\mathrm{U}} = \{1, 2, \ldots, N_{\mathrm{U}}\}$, $\mathcal{N}_{\mathrm{S}} = \{1, 2, \ldots, N_{\mathrm{S}}\}$, and $\mathcal{N}_{\mathrm{G}} = \{1, 2, \ldots, N_{\mathrm{G}}\}$, respectively.

For the SATCOM network, some communication links are not realizable due to physical restrictions. To describe the sets of the feasible links, we define sets of connectable pairs of components candidates at time k as

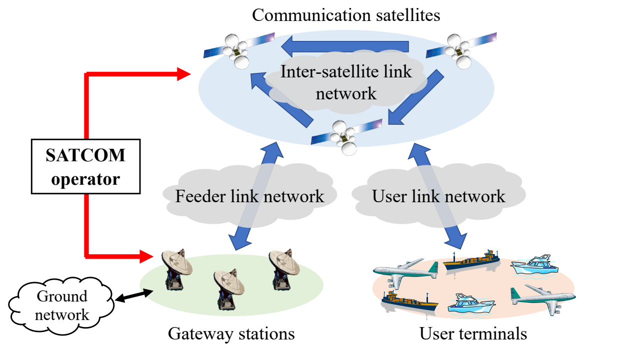

Fig. 1. Schematic of the large-scale SATCOM systems.

$$\begin{aligned}
\bar{\mathcal{C}}_k &:= \{(i,j) \in \mathcal{N}_{\mathrm{U}} \times \mathcal{N}_{\mathrm{S}} \mid i\text{-th user and } j\text{-th satellite} \\ &\qquad \text{can be connected at time } k.\}, \\
\bar{\mathcal{H}}_k &:= \{(j_1,j_2) \in \mathcal{N}_{\mathrm{S}} \times \mathcal{N}_{\mathrm{S}} \mid j_1\text{-th and } j_2\text{-th satellites} \\ &\qquad \text{can be connected at time } k.\}, \\
\bar{\mathcal{E}}_k &:= \{(j,\ell) \in \mathcal{N}_{\mathrm{S}} \times \mathcal{N}_{\mathrm{G}} \mid j\text{-th satellite and } \ell\text{-th gateway} \\ &\qquad \text{can be connected at time } k.\}.
\end{aligned}$$

These sets change over time due to time-variability in the satellite position and the propagation environment. For example, NGEO satellites rotate around the earth periodically on prescribed orbits and users cannot connect to the invisible NGEO satellites. We suppose that the SATCOM operator can obtain these sets restrictions (i.e., the sets $\bar{\mathcal{C}}_k, \bar{\mathcal{H}}_k$, and $\bar{\mathcal{E}}_k$) in advance.

2.2 Stochastic Time-Varying Communication Requests

For each time $k \geq 0$, each user can send a communication request to satellites. The amount of the communication request from the i-th user is denoted by $d_k^i \geq 0$. As described in Section 1, the stochastic and time-varying requests of the SATCOM system behave like non-stationary stochastic processes. Thus, the dynamics of the i-th user request is modeled using an auto-regressive integrated moving-average (ARIMA) model, which enables the representation of non-stationary stochastic processes Box et al. (2015). The ARIMA model of order (n_i, r_i, m_i) is represented by

$$\phi^i(B) d_{k+1}^i = \theta^i(B) w_k^i + d_{\text{const.}}^i, \tag{1}$$

where $w_k^i \in \mathbb{R}$ is an i.i.d. (independent and identically distributed) Gaussian random variable with distribution $\mathcal{N}(\mu_i, \sigma_i^2)$, $d_{\text{const.}}^i \in \mathbb{R}$ is a constant value, B is the backshift operator that works as $B^n d_k = d_{k-n}$, and $\phi^i(B) := (1 - \sum_{j=1}^{n_i} \phi_j B^j)(1-B)^{r_i}$ and $\theta^i(B) := 1 - \sum_{j=1}^{m_i} \theta_j B^j$. Here, the model in Eq. (1) is rewritten as

$$d_{k+1}^i = \tilde{\phi}^i(B) d_k^i + \theta^i(B) w_k^i + d_{\text{const.}}^i, \tag{2}$$

where $\tilde{\phi}^i(B)$ represents a coefficient function only for the past value of d_k^i.

The future value of the requests can be predicted by utilizing the dynamical model in (2) and calculating a linear sum of the time series of the past requests.

3. RESOURCE ALLOCATION AND NETWORK DESIGN PROBLEM: CC-MPC APPROACH

In this section, we formulate an optimization problem to obtain the optimal resource allocation and network structure based on the CC-MPC approach.

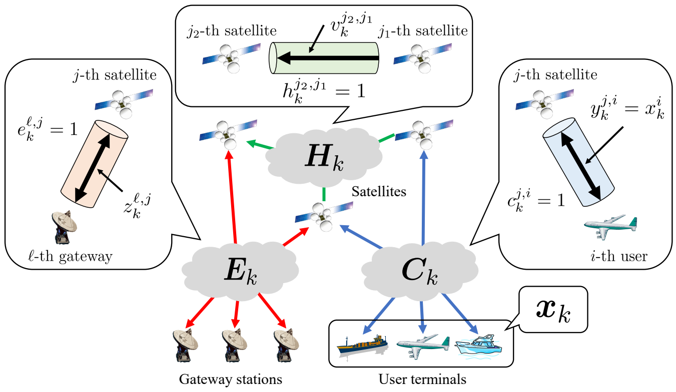

Fig. 2. Notations of the resource allocation and network structure in large-scale SATCOM systems.

3.1 Design Variables of SATCOM Operator

In this subsection, we define design variables of SATCOM network structure that the SATCOM operator plans to design. Notations appeared in this subsection are summarized in Fig. 2.

In our problem formulation, we assume that the operator of the SATCOM network can design the following two types of variables; 1) the connectivity between components (i.e., users, satellites, and gateways) and 2) the amount of data flowing over the connections. To represent the connectivities, we use the following three types of network connection matrices:

- The matrix of the bidirectional user link network at time k is denoted by $\boldsymbol{C}_k = [c_k^{j,i}]_{j,i} \in \{0,1\}^{N_\mathrm{S} \times N_\mathrm{U}}$. An entry of $\boldsymbol{C}_k$ is represented by
$$c_k^{j,i} := \begin{cases} 1, & \text{if } i\text{-th user and } j\text{-th satellite are connected,} \\ 0, & \text{otherwise.} \end{cases}$$
- The matrix of the directional inter-satellite link network at time k is denoted by $\boldsymbol{H}_k = [h_k^{j_1,j_2}]_{j_1,j_2} \in \{0,1\}^{N_\mathrm{S} \times N_\mathrm{S}}$. An entry of $\boldsymbol{H}_k$ is represented by
$$h_k^{j_2,j_1} := \begin{cases} 1, & \text{if there exists inter-satellite link} \\ & \text{from } j_1\text{-th satellite to } j_2\text{-th satellite,} \\ 0, & \text{otherwise.} \end{cases}$$
If there is no inter-satellite links, $\boldsymbol{H}_k = \boldsymbol{O}_n$ holds, where $\boldsymbol{O}_n$ denotes the $n \times n$ zero matrix.
- The matrix of the bidirectional feeder link network at time k is denoted by $\boldsymbol{E}_k = [e_k^{\ell,j}]_{\ell,j} \in \{0,1\}^{N_\mathrm{G} \times N_\mathrm{S}}$. An entry of $\boldsymbol{E}_k$ is represented by
$$e_k^{\ell,j} := \begin{cases} 1, & \text{if } j\text{-th satellite and} \\ & \ell\text{-th gateway are connected,} \\ 0, & \text{otherwise.} \end{cases}$$

By using these matrices, let $\mathcal{C}_k, \mathcal{H}_k$, and $\mathcal{E}_k$ denote sets of the connected pairs in each network defined as $\mathcal{C}_k := \{(i,j) \in \bar{\mathcal{C}}_k \mid c_k^{j,i} = 1\}, \mathcal{H}_k := \{(j_1,j_2) \in \bar{\mathcal{H}}_k \mid h_k^{j_2,j_1} = 1\}$, and $\mathcal{E}_k := \{(j,\ell) \in \bar{\mathcal{E}}_k \mid e_k^{\ell,j} = 1\}$, respectively.

In this paper, we assume that users can connect with up to one satellite in the user links and satellites can connect with up to one gateway in the feeder links, while satellites are allowed to connect with another multiple satellites in the inter-satellite links. These assumptions imply that the following constraints hold:
$$\textstyle\sum_{j:(i,j)\in\bar{\mathcal{C}}_k} c_k^{j,i} \in \{0,1\}, \quad \forall i \in \mathcal{N}_\mathrm{U}, \tag{3a}$$
$$\textstyle\sum_{\ell:(j,\ell)\in\bar{\mathcal{E}}_k} e_k^{\ell,j} \in \{0,1\}, \quad \forall j \in \mathcal{N}_\mathrm{S}. \tag{3b}$$

We next describe the amount of data flowing over the connections. The amount of bandwidth that the operator plans to allocate to i-th user at time k is denoted by
$$x_k^i \geq 0, \quad \forall i \in \mathcal{N}_\mathrm{U}. \tag{4}$$
For the N_U users, the allocated bandwidth can be written in a vector form as $\boldsymbol{x}_k = [\, x_k^1 \; x_k^2 \; \cdots \; x_k^{N_\mathrm{U}} \,]^\top \in \mathbb{R}^{N_\mathrm{U}}$. We assume that, at each time $k \geq 0$, the operator can either increase or decrease the bandwidth of users by u_k. Therefore, the dynamics of the resource allocation is given by the following discrete-time linear system:
$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{u}_k, \tag{5}$$
where $\boldsymbol{u}_k \in \mathbb{R}^{N_\mathrm{U}}$ is the control input applied at time k.

The communication data from the users flow in the communication links by utilizing the allocated bandwidth. For the user links and the feeder links, let $y_k^{j,i}$ and $z_k^{\ell,j}$ denote the amount of the flow in each link between i-th user and j-th satellite and that between j-th satellite and ℓ-th gateway at time k, respectively. For the inter-satellite links, the amount of the flow from j_1-th satellite to j_2-th satellite is represented by $v_k^{j_2,j_1}$. We then define the following data flow matrices: $\boldsymbol{Y}_k := [y_k^{j,i}]_{j,i} \in \mathbb{R}^{N_\mathrm{S}\times N_\mathrm{U}}, \boldsymbol{Z}_k := [z_k^{\ell,j}]_{\ell,j} \in \mathbb{R}^{N_\mathrm{G}\times N_\mathrm{S}}$, and $\boldsymbol{V}_k := [v_k^{j_2,j_1}]_{j_2,j_1} \in \mathbb{R}^{N_\mathrm{S}\times N_\mathrm{S}}$.

We require that the following capacity constraints for the user links and the feeder links hold:
$$\textstyle\sum_{i:(i,j)\in\mathcal{C}_k} y_k^{j,i} \leq B_\mathrm{S}^j, \quad \forall j \in \mathcal{N}_\mathrm{S}, \tag{6a}$$
$$\textstyle\sum_{j:(j,\ell)\in\mathcal{E}_k} z_k^{\ell,j} \leq B_\mathrm{G}^\ell, \quad \forall \ell \in \mathcal{N}_\mathrm{G}, \tag{6b}$$
where B_S^j and B_G^k represent the maximum bandwidths of j-th satellite and ℓ-th gateway, respectively. Furthermore, the sum of in- and out-flow in each inter-satellite link must satisfy
$$v_k^{j,\mathrm{in}} + v_k^{j,\mathrm{out}} \leq B_\mathrm{ISL}^j, \quad \forall j \in \mathcal{N}_\mathrm{S}, \tag{7}$$
where $v_k^{j,\mathrm{in}}$ and $v_k^{j,\mathrm{out}}$ represent the in- and out-flow of j-th satellite with respect to the inter-satellite links defined by
$$v_k^{j,\mathrm{in}} := \textstyle\sum_{\substack{j_1:(j_1,j)\in\mathcal{H}_k \\ j_1\neq j}} v_k^{j,j_1}, \quad v_k^{j,\mathrm{out}} := \textstyle\sum_{\substack{j_2:(j,j_2)\in\mathcal{H}_k \\ j_2\neq j}} v_k^{j_2,j},$$
and B_ISL^j represents the maximum processing bandwidth that can be used for the inter-satellite links in j-th satellite.

To ensure full and efficient use of resources, the SATCOM network is designed to conserve the in- and out-flow of each link. Thus, we require that the following conservation laws hold true:
$$x_k^i = \textstyle\sum_{j:(i,j)\in\mathcal{C}_k} y_k^{j,i}, \quad \forall i \in \mathcal{N}_\mathrm{U}, \tag{8a}$$
$$v_k^{j,\mathrm{in}} + \textstyle\sum_{i:(i,j)\in\mathcal{C}_k} y_k^{j,i} = v_k^{j,\mathrm{out}} + \textstyle\sum_{\ell:(j,\ell)\in\mathcal{E}_k} z_k^{\ell,j}, \; \forall j \in \mathcal{N}_\mathrm{S}. \tag{8b}$$

3.2 Chance Constraint of Bandwidth Loss Rate

In this subsection, a chance constraint on bandwidth loss rate is described. The bandwidth loss rate of the i-th user at time k is defined as $b_k^i := \max(0, (d_k^i - x_k^i)/d_k^i)$. The event that represents the bandwidth loss rate remaining below a threshold level $\beta_k^i \in [0,1]$ is defined as

$$\mathfrak{B}_k^i := \{\omega \in \Omega \mid b_k^i \leq \beta_k^i\}, \tag{9}$$

where ω denotes a sample from a sample space Ω.

What is crucial for the reliable operation of the SATCOM network experiencing the stochastic and time-varying requests is to ensure that the event in (9) is achieved for all the users at all times with a certain reliability level. We formulate this requirement as the *joint chance constraint* (Kall and Wallace, 1995) given by

$$\Pr\left[\bigcap\nolimits_{\tau \in \mathcal{T}_p} \mathfrak{B}_{i,k+\tau}\right] \geq 1 - \Delta^i, \quad \forall i \in \mathcal{N}_{\mathrm{U}}, \tag{10}$$

where T_p is the prediction horizon, $\mathcal{T}_p = \{1, 2, \ldots, T_p\}$ is a set of the finite-time horizon, and $\Delta^i \in (0,1]$ is the reliability level of the i-th user.

3.3 Cost Function

In practical SATCOM systems, the number of satellite handovers should be reduced because the handover is one of the burdens for both the SATCOM operator and the users. Thus, we introduce the following performance indices defined by

$$J_{\boldsymbol{C},k} := \sum\nolimits_{i \in \mathcal{N}_{\mathrm{U}}} \sum\nolimits_{j \in \mathcal{N}_{\mathrm{S}}} |c_k^{j,i} - c_{k-1}^{j,i}|, \tag{11a}$$

$$J_{\boldsymbol{H},k} := \sum\nolimits_{j_1, j_2 \in \mathcal{N}_{\mathrm{S}}} |h_k^{j_2,j_1} - h_{k-1}^{j_2,j_1}|, \tag{11b}$$

$$J_{\boldsymbol{E},k} := \sum\nolimits_{j \in \mathcal{N}_{\mathrm{S}}} \sum\nolimits_{\ell \in \mathcal{N}_{\mathrm{G}}} |e_k^{\ell,j} - e_{k-1}^{\ell,j}|, \tag{11c}$$

on $k \geq 2$. These indices directly represent the number of pairs that change the connection in the SATCOM network. Because the indices in (11) are defined on $k \geq 2$, we replace the indices with the total of the bandwidth loss rate: $\sum_{i \in \mathcal{N}_{\mathrm{U}}} b_k^i$ at time $k = 1$.

Combining the aforementioned factors, we define the cost function at time k as

$$J_k := \begin{cases} \sum_{i \in \mathcal{N}_{\mathrm{U}}} b_k^i, & k = 1, \\ w_{\boldsymbol{C}} J_{\boldsymbol{C},k} + w_{\boldsymbol{H}} J_{\boldsymbol{H},k} + w_{\boldsymbol{E}} J_{\boldsymbol{E},k}, & k \geq 2, \end{cases} \tag{12}$$

where $w_{\boldsymbol{C}} \geq 0, w_{\boldsymbol{H}} \geq 0$, and $w_{\boldsymbol{E}} \geq 0$ are the weights. The SATCOM operator designs these weights according to the management strategy.

3.4 Optimization Problem

We formulate a SATCOM network optimization problem based on the CC-MPC approach. Adopting the MPC approach, we can optimize the SATCOM network while estimating future requests of the users by using the ARIMA model in (2). To calculate the future requests of the users, we require the time series of the past requests $\{d_{k-\tau}^i\}_{0 \leq \tau \leq n_i r_i - 1}$, where n_i and r_i are the orders of the ARIMA model.

We can obtain the resource allocation and network structure of the large-scale SATCOM systems by solving the following network optimization problem at every time k.

Problem 1. (CC Network Optimization Problem).
For given $N_{\mathrm{U}}, N_{\mathrm{S}}, N_{\mathrm{G}}, B_{\mathrm{S}}^j, B_{\mathrm{ISL}}^j, B_{\mathrm{G}}^\ell$, the time series of past requests $\{d_{k-\tau}^i\}_{0 \leq \tau \leq n_i r_i - 1}$, the future values of $\{\bar{\mathcal{C}}_{k+\tau}, \bar{\mathcal{H}}_{k+\tau}, \bar{\mathcal{E}}_{k+\tau}, d_{k+\tau}^i\}_{\tau \in \mathcal{T}_p}$, and the ARIMA model in (2), solve

$$\begin{aligned} \underset{\substack{\{\boldsymbol{C}_{k+\tau}, \boldsymbol{H}_{k+\tau}, \boldsymbol{E}_{k+\tau}, \boldsymbol{Y}_{k+\tau}, \\ \boldsymbol{V}_{k+\tau}, \boldsymbol{Z}_{k+\tau}, \boldsymbol{u}_{k+\tau-1}\}_{\tau \in \mathcal{T}_p}}}{\text{minimize}} \quad & \sum\nolimits_{\tau \in \mathcal{T}_p} J_{k+\tau} \\ \text{subject to} \quad & (3)-(8), (10), \quad \forall \tau \in \mathcal{T}_p, \ i \in \mathcal{N}_{\mathrm{U}} \end{aligned}$$

Problem 1 is a mixed integer and nonlinear programming problem because $\boldsymbol{C}_k, \boldsymbol{H}_k$, and $\boldsymbol{E}_k$ are the binary matrices and the joint chance constraint in (10) is the nonlinear constraint. The SATCOM operator solves Problem 1 at every time k to obtain the optimal resource allocation and network structure. Using this result, the operator manages and controls the satellites and gateways to meet the stochastic and time-varying user requests.

4. REDUCTION TO DETERMINISTIC OPTIMIZATION PROBLEM

To solve Problem 1 efficiently, the joint chance constraint in (10) is reduced to the deterministic one through the following two stages: (Stage 1) relaxing the joint chance constraint to the individual one and (Stage 2) reducing the individual chance constraint to the deterministic one.

4.1 Stage 1: Relaxing Joint Chance Constraint

The following lemma holds for the joint chance constraint.

Lemma 1. The joint chance constraint in (10) holds if there exists $\delta_{k+\tau}^i$ ($\tau \in \mathcal{T}_p$) such that

$$\Pr\left[\mathfrak{B}_{k+\tau}^i\right] \geq 1 - \delta_{k+\tau}^i, \tag{13}$$

$$\sum\nolimits_{\tau \in \mathcal{T}_p} \delta_{k+\tau}^i \leq \Delta^i, \ \delta_{k+\tau}^i \geq 0, \tag{14}$$

hold for all $\tau \in \mathcal{T}_p$ and $i \in \mathcal{N}_{\mathrm{U}}$.

The proof of Lemma 1 is omitted here (see, e.g., Nemirovski and Shapiro (2006)).

Lemma 1 means that the joint chance constraint is divided into T_p individual chance constraints. Thus, the equations in (14) can be regarded as the reliability level Δ^i allocated to each prediction horizon as $\sum_{\tau \in \mathcal{T}_p} \delta_{k+\tau}^i$.

4.2 Stage 2: Reducing Chance Constraint

In the next stage, we reduce the chance constraint in (13) to the deterministic one.

For simplicity of notation, the time horizon index τ is neglected at first. By applying the model in (2) and $\beta_k^i \geq 0$, the inequality $b_k^i \leq \beta_k^i$ is transformed as $\bar{w}_{k-1}^i + \tilde{\phi}^i(B) d_{k-1}^i + d_{\mathrm{const.}}^i \leq x_k^i / (1 - \beta_k^i)$, where $\bar{w}_k^i := \theta^i(B) w_k^i = (1 - \sum_{j=1}^{m_i} \theta_j B^j) w_k^i$. By using this definition, $\bar{w}_k^i$ follows the Gaussian distribution with $\mathcal{N}(\bar{\mu}_i, \bar{\sigma}_i^2)$, where $\bar{\mu}_i = (1 - \sum_{j=1}^{m_i} \theta_j)\mu_i$ and $\bar{\sigma}_i^2 = (1 - \sum_{j=1}^{m_i} \theta_j^2)\sigma_i^2$. Thus, the event in (9) is equivalent to

$$\mathfrak{B}_k^i = \left\{ \omega \in \Omega \;\middle|\; \bar{w}_{k-1}^i \leq \frac{x_k^i}{(1 - \beta_k^i)} - \tilde{\phi}^i(B) d_{k-1}^i - d_{\mathrm{const.}}^i \right\}.$$

The inequality in (13) is then transformed as

$$F_{\bar{w}^i}\left(\frac{x_k^i}{(1 - \beta_k^i)} - \tilde{\phi}^i(B) d_{k-1}^i - d_{\mathrm{const.}}^i \right) \geq 1 - \delta_k^i, \tag{15}$$

where $F_{\bar{w}^i}$ represents a cumulative distribution function of $\bar{w}_k^i$. For the Gaussian with $\mathcal{N}(\bar{\mu}_i, \bar{\sigma}_i^2)$, the cumulative distribution function is obtained in an analytical form. Therefore, the inequality in (15) is transformed to

$$\frac{x_{k+\tau}^i}{(1-\beta_{k+\tau}^i)} - \tilde{\phi}^i(B)d_{k+\tau-1}^i - d_{\mathrm{const.}}^i \geq \bar{\mu}_i + \sqrt{2\bar{\sigma}_i^2}\ \mathrm{erf}^{-1}(1-2\delta_{k+\tau}^i), \quad (16)$$

where $\mathrm{erf}^{-1}(\cdot)$ is an inverse function of the error function. This is a deterministic linear inequality constraint on the $x_{k+\tau}^i$.

4.3 Relaxed Optimization Problem

From the previous subsections, Problem 1 is reduced to the following problem with the deterministic linear constraint.

Problem 2. (Relaxed Network Optimization Problem). For given $N_{\mathrm{U}}, N_{\mathrm{S}}, N_{\mathrm{G}}, B_{\mathrm{S}}^j, B_{\mathrm{ISL}}^j, B_{\mathrm{G}}^\ell$, the time series of past requests $\{d_{k-\tau}^i\}_{0\leq\tau\leq n_i r_i - 1}$, the future values of $\{\bar{\mathcal{C}}_{k+\tau}, \bar{\mathcal{H}}_{k+\tau}, \bar{\mathcal{E}}_{k+\tau}, d_{k+\tau}^i\}_{\tau\in\mathcal{T}_p}$, and the ARIMA model in (2), solve

$$\begin{aligned} \underset{\substack{\{\boldsymbol{C}_{k+\tau},\boldsymbol{H}_{k+\tau},\boldsymbol{E}_{k+\tau},\boldsymbol{Y}_{k+\tau},\\ \boldsymbol{V}_{k+\tau},\boldsymbol{Z}_{k+\tau},\boldsymbol{u}_{k+\tau-1}\}_{\tau\in\mathcal{T}_p}}}{\text{minimize}} \quad & \textstyle\sum_{\tau\in\mathcal{T}_p} J_{k+\tau} \\ \text{subject to} \quad & (3)-(8), (14), (16), \\ & \forall\tau\in\mathcal{T}_p,\ i\in\mathcal{N}_{\mathrm{U}} \end{aligned}$$

Problem 2 can be solved more efficiently than Problem 1 because the nonlinear constraint is not included. Although Problem 2 is still the mixed integer programming problem, this type of problem can be efficiently solved using the branch-and-cut algorithm, which is a heuristic algorithm (Stubbs and Mehrotra, 1999). If Problem 2 becomes infeasible at a certain time because the constraint in (16) is violated, this constraint is neglected and the total bandwidth loss rate ($\sum_{i\in\mathcal{N}_{\mathrm{U}}} b_k^i$) is directly minimized instead of the cost in (12).

5. NUMERICAL SIMULATION

In this section, we conduct a numerical simulation to verify the performance of the proposed network management methodology with the stochastic and time-varying communication requests. In this simulation, we assume that there are no inter-satellite links and satellites use different frequency bands to prevent interference between communication links. The optimization problem is described by MATLAB YALMIP (Löfberg, 2004) and solved using the Gurobi Optimizer.

5.1 Simulation Conditions

In this simulation, we assume that there are a total of 50 users and five gateways in the system. The capacity of the gateways was set to $B_{\mathrm{G}}^\ell = 500$ MHz ($\forall\ell\in\mathcal{N}_{\mathrm{G}}$). The dynamics of the user requests in (1) was set to the ARIMA model of order $(1,1,2)$. For all the users, the coefficients ϕ_1, θ_1, and θ_2 were generated from the uniform distributions on $[0.1, 0.2], [-0.2, -0.1]$, and $[-0.5, -0.4]$, respectively. Furthermore, we set $d_{\mathrm{const.}}^i = 0.5$ and $(\mu_i, \sigma_i^2) = (0.4, 5)$ for all $i\in\mathcal{N}_{\mathrm{U}}$. The simulation was conducted in $T = 50$ steps and the prediction horizon was set to $T_p = 5$.

Table 1. Simulation parameters of satellite.

Type of satellites	N_{S}	B_{S}^j [MHz] ($\forall\ell\in\mathcal{N}_{\mathrm{S}}$)	Visible period [steps]
GEO	3	500	-
NGEO group 1	24	250	5
NGEO group 2	18	250	6
(in total)	45		

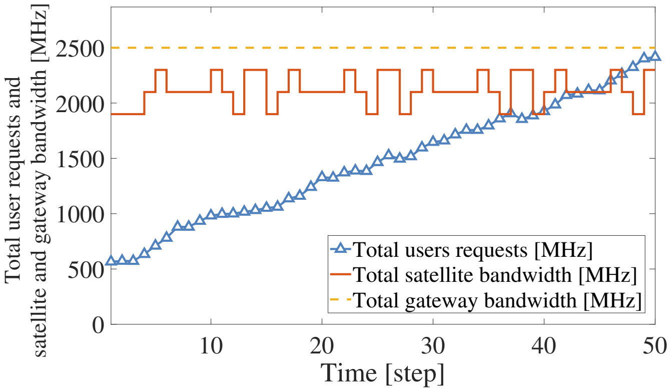

Fig. 3. Time-varying total user requests (blue line), total satellite bandwidth (red line), and total gateway bandwidth (dashed yellow line).

The simulation parameters of satellite are shown in Table 1. There are a total of 45 satellites, which includes three GEO satellites and 42 NGEO satellites divided into two groups. The satellites can connect to all users and gateways over a continuous period of visibility. The GEO satellites are visible throughout the entire simulation time interval, whereas the visibility periods of the NGEO satellites are limited to finite intervals and repeated periodically. The time-varying connectable pair candidate sets $\bar{\mathcal{C}}_k$ and $\bar{\mathcal{E}}_k$ and the total satellite capacity are calculated based on the visibility schedule. Figure 3 shows the total time-varying user requests and the total satellite and gateway bandwidth.

The reliability level Δ^i is set as $\Delta^i = 0.5, 0.3, 0.1, 0.05$, and 0.01 for all $i\in\mathcal{N}_{\mathrm{U}}$, and results for each Δ^i are compared. As shown in Fig. 4, the level Δ^i is allocated to each prediction horizon as higher reliability is required in the near future. This means that δ_{k+1}^i is the smallest among $\tau\in\mathcal{T}_p$. This setting satisfies the inequalities in (14) as $\sum_{\tau\in\mathcal{T}_p}\delta_{k+\tau}^i = \Delta^i$ ($\forall i\in\mathcal{N}_{\mathrm{U}}$). Furthermore, the bandwidth loss rate threshold is set to $\beta_k^i = 0.1$ for all the users throughout the simulation time interval. In the cost function in (12), the weights are set as $w_c = 10, w_e = 1$, and $w_h = 0$ for $k \geq 2$.

To evaluate the performance of the proposed chance-constrained network design methodology, the number of the achieved user-specified events in (9) through the simulation is defined as $\bar{N} = \sum_{i\in\mathcal{N}_{\mathrm{U}}}\sum_{k=1}^{T} N_k^i$. If $\mathfrak{B}_k^i$ is achieved, $N_k^i = 1$, otherwise, $N_k^i = 0$. An achieved rate is expressed by $\bar{N}/(N_{\mathrm{U}}T)$, which represents how many times the specified performance is achieved in all trials. Furthermore, a baseline methodology is introduced for comparison. In this baseline methodology, the chance constraint is simply replaced by the corresponding deterministic ones: $b_{k+\tau}^i \leq \beta_{k+\tau}^i, \forall i\in\mathcal{N}_{\mathrm{U}}, \tau\in\mathcal{T}_p$.

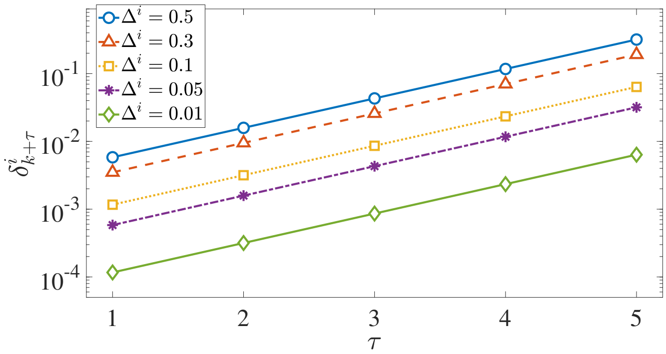

Fig. 4. Reliability allocation for each Δ^i.

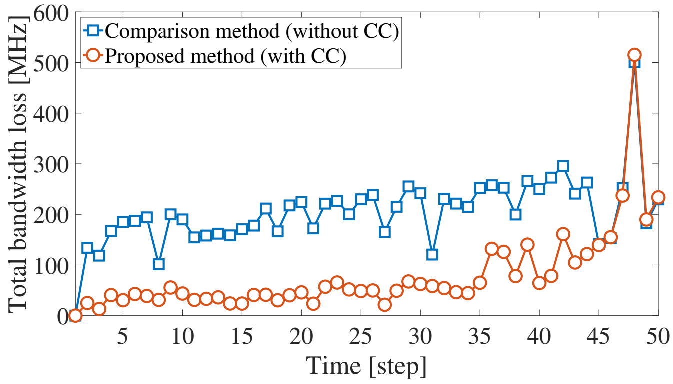

Fig. 5. Total bandwidth loss of the proposed method with $\Delta^i = 0.1$ (red line) and the comparison method (blue line).

5.2 Results

First, we compare the results of the proposed method with $\Delta^i = 0.1$ and the comparison method. In Fig. 5, the red and blue lines represent the resulting total bandwidth loss $\sum_{i\in\mathcal{N}_\mathrm{U}} \max(0, d^i_k - x^i_k)$ of the proposed method with $\Delta^i = 0.1$ and the comparison method, respectively. Note that the bandwidth loss was high after $k = 43$ for both methods because the SATCOM system could not supply enough resources to meet all the user requests. The achieved rate was 69.8% for the proposed method and 36.2% for the comparison method, and thus the proposed method outperformed the comparison method.

Next, the performance of the five different Δ^i settings are compared. Table 2 shows a comparison of the achieved rate of the events represented by $\bar{N}/(N_\mathrm{U}T)$ for each Δ^i. These results indicate that the achieved rate was increased with smaller Δ^i, which means that the specified performance achieved higher reliability. The result of the proposed method with $\Delta^i = 0.01$ performed 36.5 percentage points higher than that of the comparison method.

These results demonstrate the effectiveness of the proposed chance-constrained method in managing the stochastic and time-varying requests in the large-scale SATCOM system.

6. CONCLUSION

In this paper we proposed the chance-constrained resource and network management method for large-scale SATCOM systems with stochastic and time-varying communication requests. To efficiently solve the formulated optimization problem with the joint chance constraint on the bandwidth loss rate, the constraint is relaxed to the individual chance constraint and then reduced to the deterministic constraint. From the results of the numerical simulation, the effectiveness of the proposed method to obtain an efficient SATCOM network was verified. Furthermore, the user-specified event in (9) was achieved with higher reliability when a smaller Δ^i is adopted.

Table 2. Achieved rate of the event with Δ^i.

$\Delta^i (\forall i \in \mathcal{N}_\mathrm{U})$	Achieved rate: $\bar{N}/(N_\mathrm{U}T)$
(Comparison method)	(36.2%)
0.5 (50% reliability)	66.1%
0.3 (70% reliability)	67.0%
0.1 (90% reliability)	69.8%
0.05 (95% reliability)	70.6%
0.01 (99% reliability)	72.7%

REFERENCES

Akyildiz, I.F. and Kak, A. (2019). The Internet of Space Things/CubeSats: A ubiquitous cyber-physical system for the connected world. *Computer Networks*, 150, 134–149.

Bacco, M., Cassar, P., Colucci, M., and Gotta, A. (2018). Modeling reliable M2M/IoT traffic over random access satellite links in non-saturated conditions. *IEEE Journal on Selected Areas in Communications*, 36(5), 1042–1051.

Box, G.E.P., Jenkins, G.M., and Gregory C. Reinsel, G.M.L. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 5th edition.

del Portillo, I., G.Cameron, B., and F.Crawley, E. (2019). A technical comparison of three low earth orbit satellite constellation systems to provide global broadband. *Acta Astronautica*, 159, 123–135.

Donner, A., Kissling, C., and Hermenier, R. (2010). Satellite constellation networks for aeronautical communication: traffic modelling and link load analysis. *IET Communications*, 4(13), 1594–1606.

Kall, P. and Wallace, S.W. (1995). *Stochastic Programming*. Wiley.

Liu, J., Shi, Y., Fadlullah, Z.M., and Kato, N. (2018). Space-air-ground integrated network: A survey. *IEEE Communications Surveys & Tutorials*, 20(4), 2714–2741.

Löfberg, J. (2004). YALMIP: A toolbox for modeling and optimization in MATLAB. In *Proceedings of the 2004 IEEE International Symposium on Computer Aided Control Systems Design*, 284–289.

Maciejowski, J.M. (2000). *Predictive Control with Constraints*. Prentice Hall.

Nemirovski, A. and Shapiro, A. (2006). Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4), 969–996.

Stubbs, R.A. and Mehrotra, S. (1999). A branch-and-cut method for 0-1 mixed convex programming. *Mathematical Programming*, 86(3), 515–532.

Vasavada, Y., Gopal, R., Ravishankar, C., Zakaria, G., and BenAmmar, N. (2016). Architectures for next generation high throughput satellite systems. *International Journal of Satellite Communications and Networking*, 34(4), 523–546.