Journal of Intelligent Computing and Networking
https://www.ffspub.com/index.php/jicn/index
ISSN 3079-9228 (print)
ISSN 3107-2798 (online)
E-mail: jicn.office@ffspub.com

Article

Age of Semantics (AoS)-driven Adaptive Frame/Segment Control for Machine-centric Streaming Transmission

Ruichao Zhang[1,2], Lizhuang Tan[1,2,†], Maher Guizani[3], Wei Zhang[1,2], Hongxia Zhang[4], Peiying Zhang[4]
[1]*Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Ji'nan 250014, China*
[2]*Shandong Provincial Key Laboratory of Computing Power Internet and Service Computing, Shandong Fundamental Research Center for Computer Science, Ji'nan 250353, China*
[3]*Department of Computer Science and Engineering, University of Texas Arlington, Arlington TX 76019, USA*
[4]*Qingdao Institute of Software, College of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266580, China*
[†]*E-mail: tanlzh@sdas.org*

Received: May 15, 2026 / Revised: June 9, 2026 / Accepted: June 11, 2026 / Published online: June 16, 2026

Abstract: Machine-centric streaming transmission for inference needs to ensure both information freshness and receiver-side semantic capture rate that is available in time for inference. However, dynamic bandwidth and scene variations make these two requirements difficult to satisfy at the same time. Relying only on freshness-oriented metrics or fixed transmission modes cannot fully characterize the trade-off between freshness and semantic capture rate in Frame-by-Frame (FBF) and Segment-by-Segment (SBS) transmission. We propose Age of Semantics (AoS), which uses a computable semantic cost to jointly measure information freshness, receiver-side semantic capture rate, and the impact of scene dynamics. Based on AoS, we construct a real-time adaptive control strategy. In each control interval, the strategy compares FBF and SBS under the current network and scene states, and selects the mode with lower semantic cost. A hysteresis mechanism is used to suppress frequent switching near the decision boundary and improve online control stability. Experiments cover multiple public datasets, bandwidth levels, ablation settings, and dynamic bandwidth traces. The results show that our strategy reduces semantic transmission cost under different network and scene conditions, while maintaining stable real-time control behavior.

Keywords: Age of Information; Machine-centric Streaming Transmission; Real-time Control; Adaptive Transmission Control
https://doi.org/10.64509/jicn.22.110

1 Introduction

Streaming transmission systems are evolving from human-centric streaming to machine-centric streaming transmission, as shown in Figure 1. The former mainly targets Quality of Experience (QoE), focusing on visual clarity, stalling, and interactive experience [1], whereas the latter targets Quality of Analytics (QoA), focusing on whether the received content is timely, decodable, and usable at the inference time [2, 3]. In scenarios such as autonomous driving, UAV surveillance, intelligent inspection, and edge video analytics, the receiver mainly serves the downstream inference module [4, 5]. Therefore, pixel-level reconstruction quality is not equivalent to inference gain, and outdated information, undecodable content, or missing key semantics may weaken the reliability of analysis results [6].

In machine-centric streaming transmission, information freshness remains an important objective. Age of Information (AoI) and its extensions provide effective tools for characterizing information timeliness [7, 8]. However, from the QoA perspective, relying only on freshness-oriented metrics remains insufficient for downstream inference. For machine inference, the received content should not only be fresh, but also continuous, decodable, and available at the inference time. Therefore, freshness needs to be jointly characterized with inference-usable semantic capture rate.

Frame-by-Frame (FBF) and Segment-by-Segment (SBS) further reflect this freshness–semantic capture rate trade-off.

† Corresponding author: Lizhuang Tan
*Academic Editor: Chunxiao Jiang

FBF has a shorter waiting time and is beneficial for freshness, but it can easily lead to sparse received sequences when bandwidth is constrained or the link is congested.

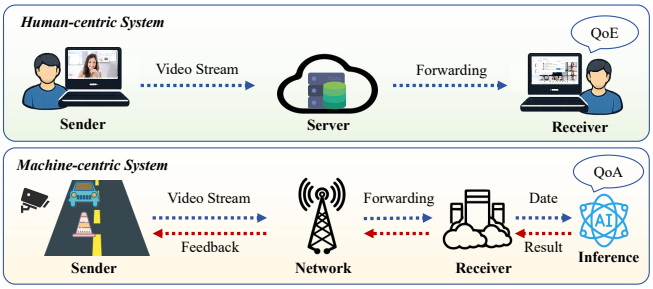

Figure 1: Differences between Human-centric and Machine-centric Streaming Transmission Systems.

SBS exploits inter-frame correlation within a segment to improve compression efficiency and preserve continuous content, but it introduces additional waiting due to segment formation, buffering, or encoding. Therefore, FBF and SBS are complementary operating points. A fixed transmission mode is difficult to maintain both low waiting time and high semantic capture rate under all network and scene conditions.

The dynamics of practical systems further require metric-control co-design. Estimated available bandwidth affects the timely transmission capability and waiting cost under different transmission modes, while scene variation intensity affects how semantic gaps degrade inference stability. Existing studies on machine-centric live streaming, task-oriented video compression, and adaptive video streaming also indicate that runtime network states, content features, and inference utility should jointly participate in transmission configuration decisions [9–11]. It should be noted that scene variation intensity only adjusts the relative weights between freshness and semantic capture rate, and does not directly determine the transmission mode. Online mode selection still needs to consider network resource states, candidate-mode costs, and historical mode states. Therefore, this paper focuses on how to realize online adaptive control for machine-centric streaming transmission under a unified semantic cost.

To address the above issues, this paper proposes the Age of Semantics (AoS) metric and the AoS-driven Adaptive Control Strategy (AAC). AoS uses Age of Semantic Information (AoSI) as the freshness-related component, and combines it with semantic capture rate and scene variation intensity. It forms a computable semantic cost for characterizing both freshness and receiver-side semantic capture rate. Furthermore, this paper transforms AoS into a log-domain semantic cost, enabling the mode-wise costs of Frame-by-Frame (FBF) and Segment-by-Segment (SBS) to be compared under the same objective. Based on this cost, AAC selects FBF or SBS online and uses a hysteresis threshold to suppress frequent switching near the decision boundary. This enables low-complexity and stable adaptive control under dynamic bandwidth and dynamic scene conditions. The main contributions of this paper are summarized as follows.

1. **Propose the Age of Semantics (AoS) metric.** AoS formulates a semantic cost for adaptive mode selection, rather than only measuring information freshness as in AoI or semantically valid updates as in AoSI. It keeps AoSI as the freshness-related term, introduces the semantic capture rate to describe receiver-side content available for inference, and uses the scene variation intensity to adjust their relative weights. This makes freshness and semantic capture rate comparable under the same cost when evaluating Frame-by-Frame (FBF) and Segment-by-Segment (SBS).
2. **Design the AoS-driven Adaptive Control Strategy (AAC).** Existing adaptive video streaming methods usually adjust bitrate, resolution, offloading, or enhancement configurations. The AoS-driven Adaptive Control Strategy (AAC) instead focuses on the choice between frame-level and segment-level transmission. It treats Frame-by-Frame (FBF) and Segment-by-Segment (SBS) as two candidate modes with different effects on freshness and semantic capture rate, compares their AoS-derived costs online, and selects the mode with lower semantic cost. A hysteresis mechanism is used to suppress frequent switching near the decision boundary.
3. **Conduct systematic experimental evaluation.** The experiments cover different bandwidth levels, dynamic bandwidth traces, multiple datasets, and ablation settings. The results validate the complementarity between FBF and SBS in semantic transmission, demonstrate the effectiveness of AoS-driven online control, and show that AAC can reduce semantic transmission cost while maintaining stable control behavior.

An earlier version of this work was published in IWCMC 2026. In the present version, we have supplemented the mathematical model, theoretical analysis, and large-scale experimental results. The remainder of this paper is organized as follows. Section 2 reviews related works on machine-centric streaming transmission, Age of Information, and adaptive transmission control, and further clarifies the motivation of this work. Section 3 establishes the system model and formulates the online mode-selection problem. Section 4 presents the Age of Semantics (AoS) metric and the AoS-driven Adaptive Control Strategy (AAC). Section 5 reports the experimental methodology, experimental results, and analysis. Section 6 concludes the paper and discusses possible future extensions.

2 Related Works and Motivation

2.1 Machine-centric Streaming Transmission

The core objective of machine-centric streaming transmission is to make video streams serve downstream machine inference, rather than human viewing experience. Xu *et al.* [2] systematically summarized the applications, architectures, and key techniques of edge video analytics, providing an important reference for system design in this direction. Following this objective, representative studies have introduced machine-centric objectives into adaptive coding, neural-enhanced video analytics streaming, and online streaming policy optimization, making the transmission process more aligned with downstream analytics requirements [12–15]. In addition, studies represented by Chen *et al.* combined edge computing with device-edge collaboration to

balance transmission overhead, computation workload, and analytics performance under resource constraints [6, 16–20].

These studies have promoted the shift of machine-centric streaming from pixel-quality optimization to inference utility, resource constraints, and device-edge collaboration. However, existing methods mainly focus on encoding configuration, compression strategy, offloading decision, or inference accuracy. They have not explicitly characterized the structural trade-off between Frame-by-Frame (FBF) and Segment-by-Segment (SBS) in terms of freshness–semantic capture rate, nor do they provide a unified semantic cost that can directly drive frame/segment-level mode selection.

2.2 Age of Information

Age of Information (AoI) is a fundamental metric for characterizing information timeliness. Kaul *et al.* [21] introduced AoI from the perspective of real-time status updating, and Yates *et al.* [7] systematically summarized its basic models, queueing mechanisms, and scheduling strategies. Since then, AoI has been extended to different communication and computing scenarios. Chiariotti *et al.* [22] proposed Query Age of Information, which evaluates information value based on freshness at the query time. Maatouk *et al.* [23] used Age of Incorrect Information to describe the cost caused by the duration of incorrect states. Meesena *et al.* [24] incorporated both transmission and processing into Age of Processed Information. For analytics-oriented video streaming, Huang *et al.* [8] further proposed Age of Semantic Information (AoSI), which introduces semantically valid updates into age-based measurement.

These metrics provide an important foundation for freshness-aware system design, but they still cannot fully characterize inference usability in machine-centric streaming transmission [25–27]. A lower information age does not necessarily mean that the received content is continuous, decodable, and usable for inference. In particular, under bandwidth limitation or link congestion, optimizing only age or semantically valid updates may favor low-waiting Frame-by-Frame (FBF) transmission, leading to sparse received sequences and missing intermediate semantics. Therefore, a semantic metric for machine inference should explicitly characterize receiver-side inference-usable semantic capture rate beyond freshness. AoI is treated as the basic timeliness metric, and AoSI is used as the freshness-related measurement based on semantically valid updates. AoS is built on this line of age-based metrics, but it is not a simple replacement of AoSI. It uses AoSI as the freshness-related component and further introduces semantic capture rate and scene variation intensity to support adaptive control between FBF and SBS.

2.3 Adaptive Transmission Control

Adaptive transmission control aims to dynamically adjust video transmission configurations according to network, computing, and content states. This field has mainly evolved along two directions. One direction focuses on semantic-aware transmission and edge video analytics, where semantic information, task requirements, or information-gathering benefits are used to adjust transmission behaviors, so that video streams can better serve downstream analytics tasks [10, 28–30]. The other direction targets edge-assisted video analytics and machine-centric streaming scenarios, where runtime resource states, contextual information, and content features are jointly used to adapt model update, visual model selection, neural coding, configuration, or collaborative learning strategies [12, 13, 19, 31–33].

These efforts indicate that runtime adaptation is essential for video analytics systems. However, they mainly adjust bitrate, resolution, transcoding level, offloading decision, or enhancement configuration. They usually focus on configuration adaptation, rather than the choice between frame-level and segment-level transmission. In contrast, this paper models FBF and SBS as two candidate modes under the freshness–semantic capture rate trade-off, and uses an AoS-derived cost to drive real-time mode selection.

2.4 Motivation

Freshness-oriented metrics are effective for characterizing information timeliness in streaming transmission, but under constrained bandwidth, they may still bias the system toward low-waiting transmission decisions. As shown in Figure 2, Frame-by-Frame (FBF) transmission has a shorter waiting time and is beneficial for freshness. However, when the link is occupied, newly generated frames continue to arrive and compete for limited transmission opportunities, causing some intermediate frames to be dropped or expired. This leads to a sparse received sequence and weakens the continuous semantic context required by downstream inference. In contrast, Segment-by-Segment (SBS) can better preserve continuous content through grouping and inter-frame compression, but it introduces additional waiting. This observation indicates that FBF and SBS are not simply superior or inferior to each other. Instead, they are two complementary modes under the freshness–semantic capture rate trade-off. Motivated by this observation, this paper aims to jointly characterize freshness and semantic capture rate under a unified semantic cost, and further designs a low-complexity online control strategy that makes stable selections between FBF and SBS according to bandwidth state and scene dynamics. The focus is not to add another video configuration, but to turn frame/segment transmission into a mode-selection problem driven by semantic cost.

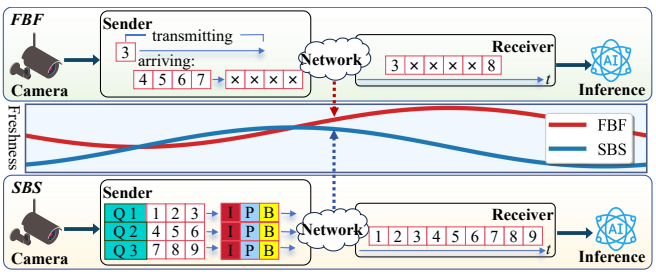

Figure 2: Two transmission model: Frame-by-Frame (FBF) and Segment-by-Segment (SBS).

3 System Model and Problem Formulation

To support the subsequent AoS metric design and online control strategy, this section formalizes the machine-centric streaming transmission system, defines the candidate transmission modes and runtime state, and formulates the resulting online mode-selection problem with a long-term optimization objective.

3.1 Notation

Table 1 summarizes the main notations used in the technical development of this paper, in order to avoid ambiguity across system modeling, metric construction, and online control. We adopt a consistent indexing convention: i denotes the input frame or video unit index, k denotes the semantic-unit index within the sliding semantic window, and t denotes the online control interval index. Unless otherwise specified, control-related states, actions, and costs are defined at the time scale of control interval t.

Table 1: Notations

Symbol	Description
x_i	Input frame or video unit
M	Candidate mode set
m_t	Selected transmission mode
$\hat{B}(t)$	Estimated available bandwidth
β_t	Scene variation intensity
s_t	Runtime state
π, Π_c	Mode-selection policy and feasible policy set
$J(t)$	Instantaneous semantic cost
$\bar{J}_\pi$	Long-term average cost under policy π
$\mathrm{AoSI}(t)$	Age of Semantic Information
$I(t)$	semantic capture rate
$\mathrm{AoS}(t)$	Age of Semantics
ε	Numerical stabilizer
$\Delta J(t)$	Mode-wise cost difference
δ_{th}	Hysteresis threshold
$W(t), K$	Sliding semantic window and window size
$q_k, s_k, a_k(t)$	Weight, encoded size, and availability indicator

3.2 Machine-centric Streaming Model

We consider a machine-centric streaming transmission system for real-time inference. The source continuously captures video content and generates a sequence of input frames or video streaming units, denoted by

$$X = \{x_i | i \geq 1\}, \tag{1}$$

where x_i represents the i-th input frame or video streaming unit. The sender encodes and packetizes the input video, which is then delivered to the receiver through a time-varying network link. The receiver performs decoding, buffering, or necessary recovery operations, and feeds the available content into the downstream inference module. The network link exhibits time-varying available bandwidth, transmission delay, and packet loss. The feedback channel provides observable information, such as ACK arrivals, receiving rate, packet losses, and inter-arrival intervals, to characterize the current transmission environment.

Unlike human-centric streaming, the effective service target of machine-centric streaming is the downstream inference process rather than pixel-level visual experience. Therefore, a transmitted unit contributes to system performance only when it arrives in time, can be decoded, and can be used by the inference module. In this paper, we use semantic capture rate to describe the portion of such useful content at the receiver, and its computable form is given in Section 4.1.2. Expired frames, undecodable frames, or content that cannot form a valid context due to missing key dependencies contributes little to this rate, even if it has already been transmitted through the network. This property implies that the system objective cannot be sufficiently described by transmission completion rate or instantaneous freshness alone. It must also account for semantic capture rate.

The sender supports two candidate transmission configurations, namely Frame-by-Frame (FBF) and Segment-by-Segment (SBS). The corresponding mode set is defined as

$$M = \{FBF, SBS\}. \tag{2}$$

In control interval t, the adopted transmission mode is denoted by $m_t \in M$. FBF uses an individual frame as the basic transmission unit. It has weaker inter-frame dependency and shorter waiting time, which is generally beneficial to information freshness. However, under bandwidth limitation or link congestion, frame-by-frame delivery forces newly generated frames to continuously compete for limited transmission opportunities, making the received sequence more likely to become sparse. *SBS* uses a segment as the basic transmission unit. It exploits inter-frame correlation to improve compression efficiency and is more likely to preserve continuous content, at the cost of additional waiting introduced by segment formation, buffering, or encoding.

Therefore, FBF and *SBS* can be viewed as two different operating points in the freshness semantic capture rate trade-off. FBF favors lower waiting time and stronger freshness, whereas *SBS* favors higher continuity and more stable semantic capture rate. Since network resources and scene content both vary over time, no fixed mode can dominate across all control intervals.

The control problem can thus be formulated as selecting an appropriate transmission mode in each control interval so as to minimize the long-term inference-oriented semantic transmission cost.

3.3 Problem Formulation

In the above system, mode selection is jointly affected by network resource conditions and content-side scene dynamics. When the available bandwidth decreases or link congestion becomes more severe, frame-by-frame transmission may prevent more newly generated frames from entering the inference pipeline in time and reduce the continuity of the received content. When scene variation becomes stronger,

the impact of missing intermediate semantics on inference stability also becomes more pronounced. In contrast, when network resources are sufficient or scene variation is weak, low-latency transmission can better exploit the advantage of freshness. We therefore formulate the problem as selecting FBF or SBS in each control interval based on the current observable state, with the objective of minimizing the long-term inference-oriented semantic transmission cost.

Let $t = 1, 2, \ldots, T$ denote the online control interval. Here, t refers to the interval in which the control decision is made, rather than the input frame index. The input frame or video unit is still indexed by i. In control interval t, the compact observable runtime state is defined as

$$s_t = (\hat{B}(t), \beta_t, m_{t-1}). \tag{3}$$

Where, $\hat{B}(t)$ denotes the estimated available bandwidth and characterizes the network resource condition. β_t denotes the current scene variation intensity and characterizes the content dynamics. m_{t-1} denotes the transmission mode adopted in the previous control interval and preserves temporal information in the mode-selection process. This state jointly captures network variation, content variation, and mode history, allowing the online mode-selection problem to be described in a unified state space.

Let π denote the mode-selection policy, and let Π_c denote the feasible policy set. The system cannot predict future bandwidth, future packet loss, or future scene variation. The mode selection in the current control interval can therefore rely only on current and historical observable information. Given state s_t, the mode selection is represented as

$$m_t = \pi(s_t), \quad m_t \in M, \quad \pi \in \Pi_c. \tag{4}$$

Where $M = \{\text{FBF}, \text{SBS}\}$ is the candidate transmission mode set. Equation 4 indicates that the system selects one transmission mode from the two candidate modes in each control interval. This selection needs to adapt to the current network resources, scene dynamics, and the mode state from the previous interval.

To keep the problem formulation general, we use $J(t)$ to denote the instantaneous semantic transmission cost in control interval t. This cost is jointly determined by the runtime state s_t and the selected mode m_t, and is written as

$$J(t) = C(s_t, m_t). \tag{5}$$

Where $C(\cdot)$ denotes the system cost function. $J(t)$ should reflect the performance loss in machine-centric streaming caused by outdated information, missing content, or inference-unusable data. At this stage, we introduce an abstract cost form to specify the system-level optimization objective. Its computable form will be derived in the metric construction.

Given policy π, the long-term average system cost is defined as

$$\bar{J}_\pi = \limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_\pi[J(t)]. \tag{6}$$

Where, $\mathbb{E}_\pi[\cdot]$ denotes the expectation under policy π, taken over network variation, content variation, and transmission randomness. The online mode-selection problem considered in this paper is formulated as

$$\pi^* = \arg\min_{\pi \in \Pi_c} \bar{J}_\pi, \tag{7}$$

subject to

$$m_t = \pi(s_t), \quad m_t \in M, \quad t = 1, 2, \ldots, T. \tag{8}$$

The above formulation captures the core control problem in machine-centric streaming transmission. The system needs to select the transmission mode online under the joint impact of dynamic network resources and dynamic scene content, while minimizing the long-term semantic transmission cost. Based on this unified problem definition, the subsequent semantic cost construction, mode-wise cost comparison, and online control rule are developed around the same objective.

4 AoS-driven Adaptive Frame/Segment Control

In this section, we further specify the mode-selection problem in machine-centric streaming transmission as an online control problem driven by semantic transmission cost. Section 3 has introduced $J(t)$ as the instantaneous semantic transmission cost in control interval t, but has not yet provided its computable form. This section instantiates $J(t)$ by constructing AoS as a computable cost that combines AoSI-based freshness and semantic capture rate. Based on this cost, we then design the AoS-driven Adaptive Control Strategy (AAC) to perform low-overhead online control between Frame-by-Frame (FBF) and Segment-by-Segment (SBS).

4.1 Age of Semantics

To instantiate the abstract cost $J(t)$ defined in Section 3, this subsection constructs Age of Semantics (AoS). AoS is a semantic cost for machine-centric streaming transmission. It keeps $AoSI(t)$ as the freshness-related component, since AoSI reflects the age of semantically valid updates. It also uses the semantic capture rate $I(t)$ to describe the content available at the receiver for inference. The scene variation intensity β_t adjusts the relative weight between these two parts. Therefore, AoS should be understood as an AoSI-based cost for adaptive FBF/SBS mode selection, rather than a simple renaming of AoSI.

4.1.1 Scene Variation Intensity

The scene variation intensity β_t characterizes the degree of scene dynamics in control interval t. In AoS, β_t only serves as a weighting factor between freshness and semantic capture rate. It does not directly determine the transmission mode. In many video analytics scenarios, stronger scene dynamics often corresponds to faster changes in object positions, occlusions, or scene context. Missing intermediate content may then have a larger impact on downstream inference. Therefore, when the scene is weakly dynamic, the system can place more emphasis on freshness; when the scene becomes

more dynamic, the system should become more sensitive to semantic capture rate.

To satisfy the low-overhead requirement of real-time control, we use compressed-domain Motion Vectors (MVs) as motion cues. The purpose is not to directly detect high-level semantic events, but to obtain a lightweight indicator of motion-related scene dynamics. In streaming video, changes that affect inference, such as object movement, occlusion, entering or leaving the view, or camera viewpoint change, are often accompanied by observable motion. MVs can be obtained from the encoded stream without running an additional inference model, which makes them suitable for online control.

Prior studies have shown that motion vectors and residual information in encoded videos can directly provide compact motion representations [34]. Specifically, the sender parses block-level MVs and aggregates the magnitude of each block motion vector into the frame-level motion intensity $\Delta_{mv}(t)$. Then, short-window smoothing is applied to $\Delta_{mv}(t)$ to suppress camera jitter and background noise. The smoothed motion intensity is normalized to [0,1] using a fixed range determined by the experimental setting and is denoted by $\bar{\Delta}_{mv}(t)$. This normalization keeps the sigmoid mapping comparable across different sequences.

In control interval t, the scene variation intensity is defined as

$$\beta_t = \frac{1}{1+\exp[-(\alpha \cdot \bar{\Delta}_{mv}(t) - \gamma)]}. \tag{9}$$

Where $\bar{\Delta}_{mv}(t)$ denotes the normalized and smoothed frame-level motion intensity, α controls the slope of the sigmoid mapping, and γ controls the midpoint. When the scene is nearly static or exhibits weak motion, β_t approaches 0. When the scene motion becomes stronger, β_t gradually approaches 1. This definition provides AoS with a low-overhead, continuous, and computable content-dynamics state.

We also note that MV-based motion intensity is not a direct measurement of high-level semantic change. Large motion may come from camera shake or background motion, while small motion may still contain task-relevant semantic changes. Short-window smoothing reduces part of this noise, but it does not remove the gap between low-level motion and high-level semantics. Therefore, β_t should be interpreted as a low-overhead scene-dynamics indicator, rather than semantic novelty or task importance. In this paper, β_t only adjusts the relative weights between freshness and semantic capture rate in AoS. The final mode selection is still determined by the mode-wise semantic cost difference $\Delta J(t)$, together with the bandwidth state and the hysteresis rule.

4.1.2 Semantic Capture Rate

The semantic capture rate $I(t)$ is the unified metric used in this paper to describe the receiver-side semantic part of AoS. It measures how much recent video content can actually contribute to downstream inference in control interval t. A higher $I(t)$ means that more useful content has arrived, has been decoded, and is available for inference. A lower $I(t)$ means that late arrival, decoding failure, link congestion, or missing dependencies have reduced the content available for inference.

Let $W(t)$ denote the sliding semantic window associated with control interval t, and let K denote the window size. The semantic units in $W(t)$ are indexed by k. In a realistic machine-centric streaming system, a semantic unit can be a frame, a decodable unit within a segment, or a video unit used by the downstream inference module. The exact granularity depends on the transmission mode and the inference task, but its role is the same. It is the basic item for computing $I(t)$.

For semantic unit k, q_k denotes its importance weight. It reflects how relevant this unit is to the downstream inference task. This is consistent with real-time video analytics studies that use spatial or temporal attention to focus processing on task-relevant content [35]. s_k denotes the encoded size of the semantic unit. It represents the transmission load of this unit, and allows $I(t)$ to account for how much encoded content has become available at the receiver. This is also consistent with edge video analytics systems where frame partition, content-aware configuration, and encoded workload affect real-time inference [36]. $a_k(t)$ denotes the availability indicator at the receiver. It is set to 1 only when semantic unit k has arrived in time, can be decoded, and can be used by the downstream inference module in control interval t. Otherwise, $a_k(t) = 0$. This interpretation is consistent with timely video analytics at the network edge, where video information is useful only when it can support inference within the required time window [37].

Based on these definitions, the semantic capture rate is defined as

$$I(t) = \frac{\sum_{k \in W(t)} q_k s_k a_k(t)}{\sum_{k \in W(t)} q_k s_k}. \tag{10}$$

In Equation (10), the numerator is the weighted encoded payload that is already available for inference within the window. The denominator is the total weighted encoded payload generated in the same window. Therefore, $I(t) \in [0,1]$. When all weighted semantic units in the window are available, $I(t)$ approaches 1. When more units become late, undecodable, or unavailable for inference, $I(t)$ decreases. In this sense, $I(t)$ gives a direct receiver-side measure of semantic capture rate for AoS computation.

4.1.3 AoS

After obtaining β_t and $I(t)$, AoS integrates freshness and semantic capture rate into a unified semantic cost. In this formulation, $AoSI(t)$ is the freshness-related term, $I(t)$ represents the semantic capture rate, and β_t controls their relative weights according to the current scene variation intensity.

We adopt a multiplicative form because $AoSI(t)$ and $I(t)$ have opposite cost directions. A larger $AoSI(t)$ indicates older semantic information and should increase the cost, while a larger $I(t)$ indicates that more content is available for inference and should reduce the cost. Therefore, $AoSI(t)$ is placed in the numerator, and $(I(t)+\varepsilon)$ is placed in the denominator. The exponents $(1-\beta_t)$ and β_t further provide scene-dependent weighting between the two parts. AoS is defined as

$$AoS(t) \triangleq \frac{(AoSI(t))^{1-\beta_t}}{(I(t)+\varepsilon)^{\beta_t}}. \tag{11}$$

Where ε is a numerical stabilizer used in AoS computation to avoid numerical instability when $I(t)$ approaches 0. We use

$\varepsilon = 10^{-4}$ by default. $AoS(t)$ is a semantic cost, and a smaller value indicates a better overall state between freshness and semantic capture rate.

Equation (11) has clear boundary implications. When β_t approaches 0, $AoS(t)$ approximately degenerates to $AoSI(t)$, and the cost is mainly determined by the freshness-related term. When β_t approaches 1, $AoS(t)$ becomes more sensitive to the decline of $I(t)$, which means that missing semantics is penalized more strongly under highly dynamic scenes. As shown in Figure 3, increasing β_t makes AoS more sensitive to the decrease in semantic capture rate, which is consistent with the need to keep useful content available for inference in dynamic scenes. This boundary behavior shows the role of the multiplicative form. It combines two quantities with opposite cost directions, while allowing β_t to continuously adjust the sensitivity of AoS to freshness and semantic capture rate.

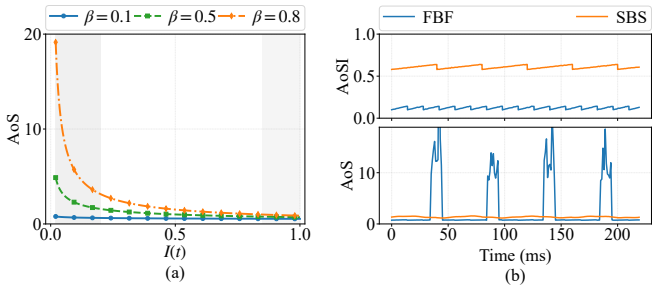

Figure 3: Impact of β_t on AoS.

For mode-wise cost comparison and online control, we transform AoS into a log-domain semantic cost, and instantiate the instantaneous semantic transmission cost $J(t)$ in Section 3. The logarithm is a monotonic transformation, so comparing $J(t) = \ln AoS(t)$ gives the same mode preference as comparing $AoS(t)$. Therefore, the log-domain form does not change the control decision between FBF and SBS. Its role is to rewrite the multiplicative cost into an additive form, so that the two loss terms can be interpreted and compared more directly. We define

$$J(t) = \ln AoS(t) = (1-\beta_t)\ln AoSI(t) - \beta_t \ln(I(t)+\varepsilon). \quad (12)$$

This form converts the multiplicative trade-off into an additive decomposition. The first term, $(1-\beta_t)\ln AoSI(t)$, represents the freshness-related loss. When $AoSI(t)$ increases, semantic information becomes older and this term increases. The second term, $(-\beta_t \ln(I(t)+\varepsilon))$, represents the loss caused by a low semantic capture rate. When $I(t)$ decreases, $\ln(I(t)+\varepsilon)$ becomes smaller, and this negative log term increases the cost. Therefore, $J(t)$ penalizes both outdated semantic information and insufficient semantic capture rate. For FBF and SBS, substituting their mode-dependent $AoSI(t)$ and $I(t)$ into Equation (12) gives $J_{FBF}(t)$ and $J_{SBS}(t)$. This makes the later cost difference $\Delta J(t) = J_{SBS}(t) - J_{FBF}(t)$ easy to interpret: a negative value favors SBS, and a positive value favors FBF.

4.2 AoS-driven Adaptive Control Strategy

Based on the above AoS cost, we design the AoS-driven Adaptive Control Strategy (AAC). The key idea of AAC is to compare the mode-wise costs of FBF and SBS in each control interval according to the network resource condition and scene dynamics, and to suppress frequent switching through a hysteresis mechanism when the two costs are close. AAC focuses on the choice between frame-level FBF and segment-level SBS, rather than general bitrate or resolution adaptation. It does not rely on a single state variable, but selects the transmission mode online using the unified AoS cost.

4.2.1 Decision Space

For a given state s_t, AAC estimates the log-domain AoS costs under the two candidate modes, denoted by $J_{\mathrm{FBF}}(t)$ and $J_{\mathrm{SBS}}(t)$, respectively. They are obtained by substituting the mode-dependent $\mathrm{AoSI}(t)$ and $I(t)$ into Equation (12). The Cost Estimator obtains these two costs in a mode-wise manner. For each candidate mode $m \in M = \{FBF, SBS\}$, it first estimates the mode-dependent freshness term $AoSI_m(t)$ and semantic capture rate $I_m(t)$ under the current bandwidth $B(t)$ and the characteristics of mode m.

$B(t)$ affects how many video units can be delivered in time and how much waiting is introduced. The candidate mode determines the transmission behavior. FBF usually has a shorter waiting time, but under low bandwidth it may lead to a sparse received sequence. SBS can preserve more continuous content through segment coding, but it introduces segment formation and buffering delay. Therefore, $AoSI_m(t)$ reflects the expected freshness-related cost of mode m, and $I_m(t)$ reflects the expected receiver-side semantic capture rate under that mode.

The scene variation intensity β_t does not directly select the mode. Instead, after $AoSI_m(t)$ and $I_m(t)$ are estimated for FBF and SBS, β_t weights the freshness-related term and the semantic-capture-rate term through Equation (12). In this way, the Cost Estimator obtains $J_{FBF}(t)$ and $J_{SBS}(t)$, and the two costs are then used to compute $\Delta J(t)$ for online mode selection.

To explain the mode preference on the $(B(t)-\beta_t)$ plane, we normalize the estimated available bandwidth as

$$B_n(t) = \min\left(\frac{B(t)}{B_{\max}}, 1\right). \quad (13)$$

where $B_n(t)$ denotes the normalized bandwidth, and $B_{\max}$ denotes the bandwidth upper bound. In our experimental setting, $B_{\max} = 8$ Mbps. This normalization is only used for illustrating the decision space in Figure 4, and does not change the definition of $B(t)$ as the estimated available bandwidth.

In control interval t, the mode-wise cost difference between FBF and SBS is defined as

$$\Delta J(t) = J_{\mathrm{SBS}}(t) - J_{\mathrm{FBF}}(t). \quad (14)$$

When $\Delta J(t) < 0$, the semantic cost of SBS is lower than that of FBF, indicating that SBS is more advantageous. When $\Delta J(t) > 0$, the semantic cost of FBF is lower than that of SBS. Since $B(t)$ affects the amount of content that can be delivered in time and the waiting time under the two modes, while β_t adjusts the relative weights between freshness and semantic capture rate in AoS, these two quantities jointly determine the relative costs of the two modes.

As shown in Figure 4, SBS tends to be more advantageous in the region with low normalized bandwidth and high β_t, because preserving continuous semantics becomes more important for reducing AoS in this case. In contrast, FBF tends to be more advantageous when the normalized bandwidth is high or β_t is low, because lower waiting time and stronger freshness can bring more benefit. Near the boundary, the costs of FBF and SBS become close, and $\Delta J(t)$ can be affected by short-term estimation fluctuations. Therefore, a hysteresis threshold is introduced in the subsequent mode control to avoid frequent switching near the boundary.

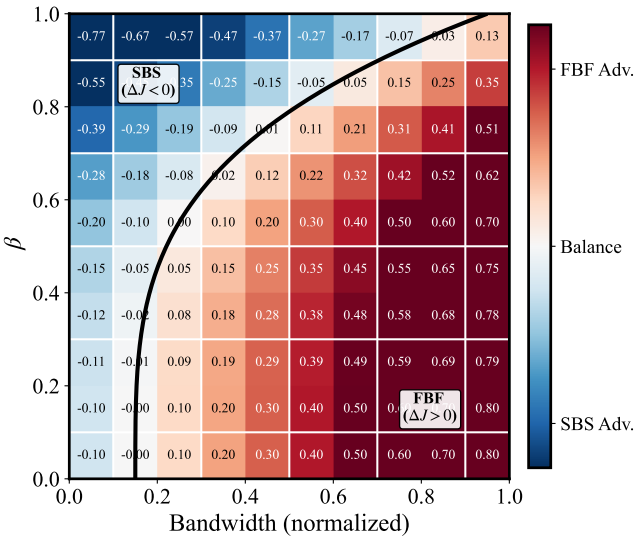

Figure 4: Decision space.

4.2.2 System Framework

As shown in Figure 5, AAC consists of a lower-layer transmission loop and an upper-layer control plane. The lower-layer transmission loop is responsible for encoding, transmission, decoding and recovery, and inference execution. The upper-layer control plane performs runtime state observation, mode-wise cost estimation, mode decision, and mode switching. This layered structure allows the video stream to keep running, while AAC updates the transmission mode at the control-interval scale.

In each control interval t, AAC works as follows. The transmission loop first sends video units using the previous mode m_{t-1}. During this process, the receiver returns ACKs and other feedback information, such as receiving rate, packet loss, and inter-arrival intervals. The State Observer uses this feedback to update the estimated bandwidth $B(t)$, and obtains β_t from compressed-domain motion cues. Together with m_{t-1}, these variables form the current runtime state. The Cost Estimator then estimates the two mode-wise costs $J_{FBF}(t)$ and $J_{SBS}(t)$ under the current state. The Decision Logic computes $\Delta J(t) = J_{SBS}(t) - J_{FBF}(t)$, compares it with the hysteresis threshold δ_{th}, and generates the new mode m_t. Finally, the Mode Switching module applies m_t to the sender-side encoder, so that the selected FBF or SBS configuration is used in the next control interval.

This feedback framework is consistent with recent edge-assisted video analytics and adaptive video streaming systems, where runtime context observations are used to drive configuration updates under dynamic network and computing environments [29, 30]. In implementation, the Cost Estimator can obtain $J_{FBF}(t)$ and $J_{SBS}(t)$ through a precomputed lookup table indexed by $B(t)$ and β_t, or through a lightweight cost evaluator. In this way, AAC does not need to run the full cost model repeatedly during streaming. Each interval only requires state update, cost lookup or estimation, cost-difference computation, threshold comparison, and encoder-mode update. Therefore, the deployment process is a closed online loop. Feedback updates the state, the state updates the costs, the costs determine m_t, and m_t is applied to the encoder for the next interval.

4.2.3 Mode Control Rule

The mode control rule of AAC uses the AoS-derived cost difference as the unified criterion. Given $\Delta J(t) = J_{\mathrm{SBS}}(t) - J_{\mathrm{FBF}}(t)$, if $\Delta J(t) < 0$, then SBS has a lower cost. If $\Delta J(t) > 0$, then FBF has a lower cost. To avoid frequent switching caused by estimation errors or short-term fluctuations when $\Delta J(t)$ is close to 0, AAC introduces a hysteresis threshold δ_{th}. Based on the above mode-wise cost difference and hysteresis mechanism, AAC performs mode selection in each control interval according to Algorithm 1.

Algorithm 1 AoS-driven Adaptive Frame/Segment Control

Require: $B(t), \beta_t, m_{t-1}, \delta_{\mathrm{th}}$, cost evaluator or lookup table.
Ensure: selected transmission mode m_t.

```
Obtain J_FBF(t) and J_SBS(t) under the current state.
Compute ΔJ(t) = J_SBS(t) − J_FBF(t).
if ΔJ(t) < −δ_th then
    m_t ← SBS
else if ΔJ(t) > δ_th then
    m_t ← FBF
else
    m_t ← m_{t−1}
end if
Apply m_t to the encoder for the next control interval.
```

Equivalently, the mode-selection rule can be written in the following compact piecewise form

$$m_t = \begin{cases} \mathrm{SBS}, & \Delta J(t) < -\delta_{\mathrm{th}}, \\ \mathrm{FBF}, & \Delta J(t) > \delta_{\mathrm{th}}, \\ m_{t-1}, & |\Delta J(t)| \le \delta_{\mathrm{th}}. \end{cases} \tag{15}$$

Equation (15) indicates that the system selects SBS when SBS has a sufficiently clear cost advantage over FBF. It selects FBF when the cost advantage of FBF exceeds the threshold. When the cost difference between the two modes lies within the hysteresis interval, the system keeps the previous mode m_{t-1}. This rule prevents AAC from switching frequently near the mode boundary due to short-term estimation fluctuations, thereby improving online control stability.

5 Experimental Evaluation

5.1 Experimental Methodology

This section evaluates the effectiveness of AoS and AAC in machine-centric streaming transmission. The baselines

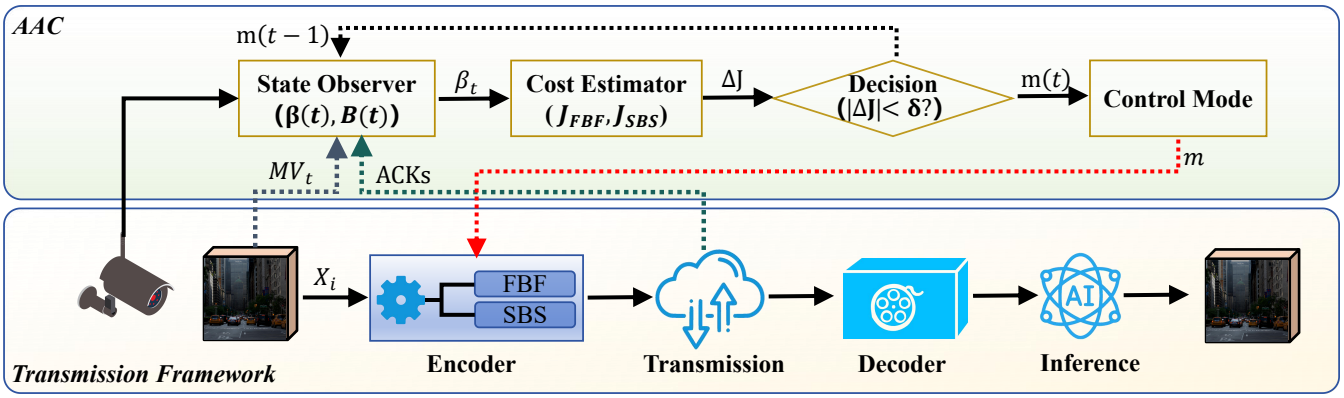

Figure 5: Framework of AAC.

include fixed-mode FBF, fixed-mode SBS, and an AoSI-oriented baseline. The ablation variants remove β_t, $B(t)$, and the hysteresis mechanism, respectively. The evaluation metrics include $I(t)$, AoS(t), average AoS, and switching count. Controlled simulations are used to explain parameter sensitivity and dynamic robustness, while experiments on public datasets provide the main validation. The main experimental settings are summarized in Table 2.

Table 2: Experimental Parameters

Parameter	Setting
Datasets	MOT17,MOT20[38], SeaDronesSee[39], VisDrone[40]
Inference model	YOLOv10 [41]
Input resolution	640×640
Confidence threshold	0.25
NMS IoU threshold	0.45
Bandwidth range	0–8 Mbps
Bandwidth upper bound	$B_{\max} = 8$ Mbps
Numerical stabilizer	$\varepsilon = 10^{-4}$
Hysteresis threshold	$\delta_{\text{th}} = 0.05$
Controlled simulation seed	42
Controlled simulation runs	30 independent runs

5.2 Experimental Results

This section analyzes the experimental results from five aspects: mode complementarity, online control behavior, overall semantic cost, cross-scenario consistency, and control stability. Overall, FBF and SBS exhibit clear complementarity under different bandwidth and scene dynamics. AoS integrates freshness and semantic capture rate into a unified evaluation metric, while AAC can make stable selections between the two candidate modes according to runtime states and mode-wise cost differences.

Mode complementarity. We first examine the difference between FBF and SBS in terms of semantic capture rate. As shown in Figure 6, $I(t)$ of FBF is significantly lower in the low-bandwidth region. This indicates that frame-by-frame transmission, although having a lower waiting time, may cause newly generated frames to continuously compete for limited transmission opportunities under constrained links, resulting in sparse received sequences. In contrast, SBS maintains a more stable $I(t)$ under low bandwidth, which is consistent with its mechanism of preserving continuous content by exploiting inter-frame correlation within a segment. As bandwidth increases, $I(t)$ of FBF rises rapidly and surpasses SBS in the medium- and high-bandwidth regions. This shows that when network resources are sufficient, low-latency transmission can better exploit the benefits of freshness and timely availability. These results indicate that FBF and SBS do not have a globally fixed superiority relationship. Instead, they represent two complementary operating points in the freshness–semantic capture rate trade-off, providing direct motivation for adaptive mode control.

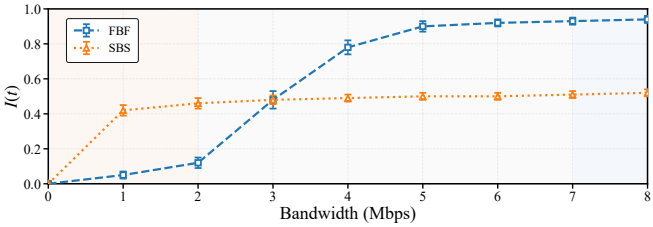

Figure 6: Semantic complementarity of FBF and SBS.

Real-time control behavior. As shown in Figure 7, it further illustrates the online behavior of AAC under dynamic bandwidth and scene variation. The bandwidth trajectory $B(t)$, the scene variation intensity β_t, and the mode-wise cost difference $\Delta J(t)$ jointly affect the mode selection. When the bandwidth is low and β_t is high, semantic continuity becomes more important for reducing AoS, and AAC tends to select SBS. When the bandwidth increases or the freshness advantage becomes more pronounced, AAC tends to select FBF. The AoS curves on the right show that the cost trajectory of AAC stays close to the more suitable fixed mode at different stages, rather than being constrained by a single configuration throughout the transmission process. Meanwhile, when $\Delta J(t)$ approaches the boundary region around zero, the hysteresis threshold suppresses frequent switching caused by short-term estimation perturbations. These results validate the control logic of Algorithm 1 in Section 4. AAC is not triggered by a single bandwidth or scene variable, but performs online selection based on the AoS-derived mode-wise cost difference.

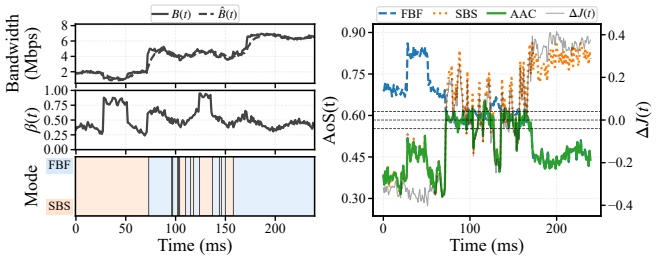

Figure 7: Real time control under temporal variations.

Bandwidth-wise performance. As shown in Figure 8, it compares the average AoS of FBF, SBS, the AoSI-oriented baseline, and AAC across bandwidth levels. In the low-bandwidth region, FBF yields the highest AoS, mainly because limited semantic capture rate makes it difficult for the receiver to form continuous and content available for inference. SBS is more stable under low bandwidth, but as bandwidth further increases, the waiting cost introduced by segment formation and buffering limits its ability to further reduce AoS. The AoSI-oriented baseline reflects the effect of freshness-related age, but without explicitly modeling semantic capture rate, it still cannot avoid the cost caused by insufficient semantic capture rate in some bandwidth regions. In contrast, AAC maintains a lower AoS across the entire bandwidth range and stays close to the currently more suitable mode. This indicates that AoS-driven control can avoid the performance degradation of fixed modes in mismatched bandwidth regions, while jointly exploiting the semantic-completeness benefit under low bandwidth and the freshness benefit under high bandwidth.

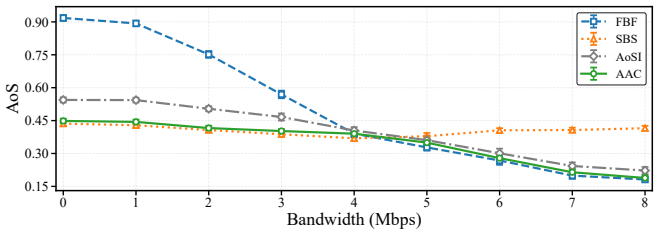

Figure 8: Performance across bandwidth.

Cross-scenario consistency. As shown in the left panel of Figure 9, AAC maintains a relatively low average AoS across different datasets. MOT17 and MOT20 mainly cover urban pedestrians, dense occlusions, and continuous motion, whereas SeaDronesSee and VisDrone include maritime targets, UAV viewpoints, small objects, and complex background variations. Although these datasets differ considerably in object scale, motion pattern, and scene background, AAC maintains a relatively low AoS across multiple scenarios and shows a consistent trend. This result indicates that the joint characterization of scene variation intensity and semantic capture rate in AoS is not limited to a single dataset setting, and that AAC can maintain stable mode-selection behavior under different scene contents.

Component necessity. As shown in the right panel of Figure 9, it reports the ablation results of the key components. Compared with Full AAC, w/o β_t removes the scene variation intensity and therefore cannot adjust the relative weights between freshness and semantic capture rate according to content dynamics. This makes the cost evaluation more likely to become mismatched in dynamic scenes. w/o $B(t)$ removes the bandwidth-state input, making it difficult for the controller to promptly reflect the impact of network resource variation on the two transmission modes. w/o hysteresis weakens switching stability near the decision boundary, making the system more vulnerable to short-term estimation fluctuations. Full AAC achieves the lowest or near-lowest average AoS, indicating that β_t, $B(t)$, and the hysteresis mechanism are not separate new algorithms, but necessary components that make AoS-driven online control both adaptive and stable.

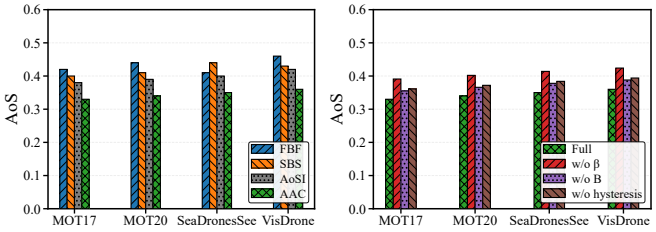

Figure 9: Public dataset results and ablation results.

Hysteresis trade-off. To further analyze the impact of the hysteresis threshold, we compare the average semantic cost $J(t)$ and switching count under different δ_{th} settings, as shown in Figure 10. When $\delta_{\text{th}} = 0$, the controller reacts frequently to short-term fluctuations in $\Delta J(t)$, resulting in a high switching count. As δ_{th} increases, the number of switches decreases rapidly, indicating that the hysteresis mechanism can effectively suppress unnecessary mode switching near the decision boundary. Meanwhile, the average semantic cost $J(t)$ remains low around $\delta_{\text{th}} = 0.05$. Further increasing the threshold brings limited additional reduction in switching count, while an overly strong hysteresis effect may delay the response to real state changes and increase the cost. Therefore, $\delta_{\text{th}} = 0.05$ is more suitable as a balanced operating point between responsiveness and switching stability, rather than being interpreted as a universally significant optimum across all settings.

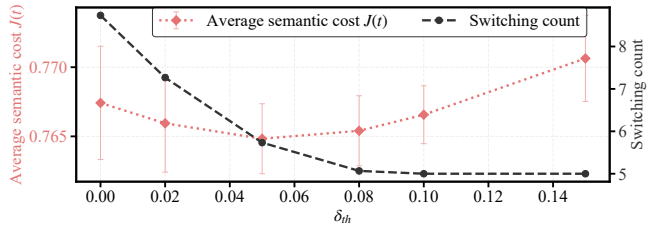

Figure 10: Sensitivity to Hysteresis Threshold.

Trace robustness. Finally, additional controlled experiments evaluate the robustness of AAC under three representative dynamic bandwidth traces: stable, bursty, and gradual. The results are shown in Figure 11. Under the stable trace, the average semantic cost $J(t)$ of AAC is close to that of the best baseline, indicating that adaptive control does not introduce noticeable extra cost when the environment is relatively stable. Under the bursty trace, the cost of fixed FBF increases significantly, reflecting the higher sensitivity of frame-by-frame transmission to sudden bandwidth drops. AAC achieves the lowest average semantic cost, showing that it can make more appropriate online selections between FBF and SBS according to state changes. Under the gradual trace,

AAC also maintains the lowest or near-lowest cost, indicating that it can smoothly adapt to slowly changing network conditions. In terms of switching count, fixed FBF and fixed SBS naturally do not switch modes, so the more meaningful comparison is with the AoSI-oriented adaptive baseline. Under dynamic traces, AAC has a switching count lower than or close to that baseline, further showing that the hysteresis mechanism helps reduce unnecessary switching and improve online control stability.

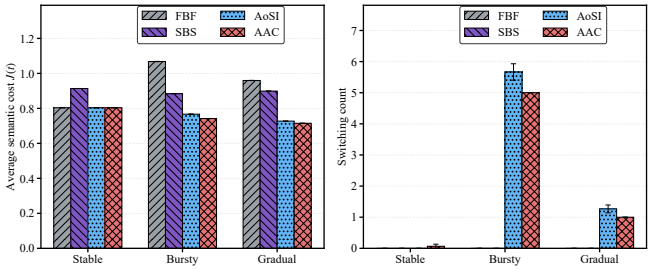

Figure 11: Robustness under Dynamic Bandwidth Traces.

Taken together, these results validate the effectiveness of AoS and AAC from multiple perspectives. The complementarity between FBF and SBS shows that a single fixed mode cannot cover all network and scene states. The online behavior analysis demonstrates that AAC can perform dynamic mode selection according to the AoS-derived cost difference. The bandwidth-wise and cross-scenario results indicate that AAC can maintain a lower semantic transmission cost under different conditions. The ablation, threshold sensitivity, and dynamic-trace analyses further show that β_t, $B(t)$, and the hysteresis mechanism are necessary for improving both adaptivity and control stability. These results jointly support the core argument of this paper: machine-centric streaming transmission should consider freshness, semantic capture rate, and online control in a unified manner, rather than relying only on a single freshness metric or a fixed transmission configuration.

6 Conclusion

AoI and AoSI provide useful freshness-oriented measurements for streaming systems. However, machine-centric streaming transmission also needs to describe whether enough content is available for inference at the receiver. Based on this observation, this paper proposed the Age of Semantics (AoS) metric and the AoS-driven Adaptive Control Strategy (AAC). AoS uses AoSI as the freshness-related component, and combines it with semantic capture rate and scene variation intensity. Based on this metric, AAC performs low-complexity online selection between FBF and SBS according to the estimated bandwidth $B(t)$, the scene variation intensity β_t, and the mode-wise cost difference $\Delta J(t)$, while using a hysteresis threshold to suppress frequent switching. Experimental results show that FBF and SBS are complementary under different bandwidth and scene conditions. AAC maintains a lower semantic transmission cost under dynamic conditions and demonstrates stable control behavior across bandwidth levels, scenarios, and ablation settings.

Future work will further extend the applicability of AoS-driven control. First, the current two-mode setting with FBF and SBS can be extended to richer encoding and transmission configurations, such as different segment lengths, resolutions, frame rates, and semantic-priority transmission strategies. Second, more systematic deployment validation can be conducted with real edge inference pipelines, real network traces, and multi-task inference scenarios. Finally, β_t, δ_{th}, and the cost estimation mechanism can be further adapted online to improve generalization across different cameras, network conditions, and inference tasks. Overall, this paper shows that machine-centric streaming transmission should jointly consider freshness, semantic capture rate, and online control, rather than relying only on a single freshness metric or a fixed transmission configuration.

Funding

This work was supported by the National Key R&D Program of China under Grant No.2024YFB2907000, the Shandong Provincial Natural Science Foundation under Grant No.ZR2023LZH017, the National Natural Science Foundation of China under Grant No.62471493, the Industry–University Research Innovation Foundation of the Ministry of Education of China under Grant No.2024IT019, and the Pilot Project for Integrated Innovation of Science, Education and Industry of Qilu University of Technology (Shandong Academy of Sciences) under Grant No.2025ZDZX01.

Author Contributions

Conceptualization, R.Z. and L.T.; methodology, R.Z.; software, R.Z.; validation, L.T.; formal analysis, W.Z. and R.Z.; writing—original draft preparation, H.Z. and R.Z.; writing—review and editing, M.G.; visualization, R.Z.; supervision, P.Z.; project administration, P.Z.; funding acquisition, P.Z. All authors have read and agreed to the published version of the manuscript.

Conflict of Interest

All the authors declare that they have no conflict of interest.

Data Available

Not applicable.

Ethical Approval

Not applicable.

Acknowledgements

An earlier version of this work was published in IWCMC 2026. In the present version, we have supplemented the mathematical model, theoretical analysis, and large-scale experimental results.

References

[1] Shi, W., Li, Q., Yu, Q., Wang, F., Shen, G., Jiang, Y.,

Xu, Y., Ma, L., Muntean, G.-M.: A Survey on Intelligent Solutions for Increased Video Delivery Quality in Cloud–Edge–End Networks. IEEE Communications Surveys & Tutorials **27**(2), 1363–1394 (2025). https://doi.org/10.1109/COMST.2024.3427360

[2] Xu, R., Razavi, S., Zheng, R.: Edge Video Analytics: A Survey on Applications, Systems and Enabling Techniques. IEEE Communications Surveys & Tutorials **25**(4), 2951–2982 (2023). https://doi.org/10.1109/COMST.2023.3323091

[3] Shao, J., Zhang, X., Zhang, J.: Task-Oriented Communication for Edge Video Analytics. IEEE Transactions on Wireless Communications **23**(5), 4141–4154 (2024). https://doi.org/10.1109/TWC.2023.3314888

[4] Zhang, Y., Zhang, W., Du, H., Yan, C., Liu, L., Zheng, Q.: FHVAC: Feature-Level Hybrid Video Adaptive Configuration for Machine-Centric Live Streaming. IEEE Transactions on Parallel and Distributed Systems **35**(5), 780–795 (2024). https://doi.org/10.1109/TPDS.2024.3372046

[5] Xiao, X., Zuo, Y., Yan, M., Wang, W., He, J., Zhang, Q.: Task-Oriented Video Compressive Streaming for Real-Time Semantic Segmentation. IEEE Transactions on Mobile Computing **23**(12), 14396–14413 (2024). https://doi.org/10.1109/TMC.2024.3,446185

[6] Chen, S., Yin, J., Zhong, R., Liu, F.: DeVA: An Edge-Assisted Video Analytics Framework for Depth Estimation. IEEE Transactions on Mobile Computing **24**(12), 13177–13190 (2025). https://doi.org/10.1109/TMC.2025.3588864

[7] Yates, R.D., Sun, Y., Brown, D.R., Kaul, S.K., Modiano, E., Ulukus, S.: Age of Information: An Introduction and Survey. IEEE Journal on Selected Areas in Communications **39**(5), 1183–1210 (2021). https://doi.org/10.1109/JSAC.2021.3065072

[8] Huang, Z., Wu, W., Wu, K., Gao, G., Wang, J.: Minimizing Age of Semantic Information for Analytics-Oriented Video Streaming Systems. IEEE Transactions on Mobile Computing **24**(12), 12885–12902 (2025). https://doi.org/10.1109/TMC.2025.3588474

[9] Feng, D., Wang, L., Chen, S., Tung, L., Liu, F.: X-Stream: A Flexible, Adaptive Video Transformer for Privacy-Preserving Video Stream Analytics. In Proceedings of IEEE INFOCOM 2024 - IEEE Conference on Computer Communications, pp. 1–10 (2024). https://doi.org/10.1109/INFOCOM52122.2024.10621341

[10] Zhang, X., Xu, H., Zou, L., Duan, J., Wu, C., Xue, Y., Chen, Z., Chen, X.: Rosevin: Employing Resource- and Rate-Adaptive Edge Super-Resolution for Video Streaming. In Proceedings of IEEE INFOCOM 2024 - IEEE Conference on Computer Communications, pp. 491–500 (2024). https://doi.org/10.1109/INFOCOM52122.2024.10621104

[11] He, Y., Yang, P., Qin, T., Hou, J., Zhang, N.: Joint Encoding and Enhancement for Low-Light Video Analytics in Mobile Edge Networks. IEEE Transactions on Mobile Computing **24**(4), 3330–3345 (2025). https://doi.org/10.1109/TMC.2024.3514214

[12] Zhu, A., Zhang, S., Cheng, K., Shi, X., Qian, Z., Lu, S.: AdaStreamer: Machine-Centric High-Accuracy Multi-Video Analytics with Adaptive Neural Codecs. In Proceedings of IEEE INFOCOM 2024 - IEEE Conference on Computer Communications, pp. 1161–1170 (2024). https://doi.org/10.1109/INFOCOM52122.2024.10621074

[13] Cen, S., Zhang, M., Zhu, Y., Liu, J.: AdaDSR: Adaptive Configuration Optimization for Neural Enhanced Video Analytics Streaming. IEEE Internet of Things Journal **11**(7), 11919–11929 (2024). https://doi.org/10.1109/JIOT.2023.3331699

[14] Li, Z., Zhang, M., Zhu, Y.: OAVS: Efficient Online Learning of Streaming Policies for Drone-sourced Live Video Analytics. In 2024 IEEE/ACM 32nd International Symposium on Quality of Service (IWQoS), pp. 1–10 (2024). https://doi.org/10.1109/IWQoS61813.2024.10682870

[15] Zhang, Y., Zhang, W., Yuan, M., Xu, L., Yan, C., Gong, T., Du, H.: Lightweight Configuration Adaptation With Multi-Teacher Reinforcement Learning for Live Video Analytics. IEEE Transactions on Mobile Computing **24**(5), 4466–4480 (2025). https://doi.org/10.1109/TMC.2025.3526359

[16] Li, X., Bi, S., Wang, S., Li, X., Zhang, Y.-J.A.: Digital semantic device-edge co-inference with task-oriented ARQ. IEEE Transactions on Vehicular Technology **73**(9), 13986–13990 (2024). https://doi.org/10.1109/TVT.2024.3390213

[17] Peng, Y., Xiang, L., Yang, K., Wang, K., Debbah, M.: Semantic Communications With Computer Vision Sensing for Edge Video Transmission. IEEE Transactions on Mobile Computing **25**(6), 7988–8001 (2026). https://doi.org/10.1109/TMC.2025.3646710

[18] Gao, G., Dong, Y., Wang, R., Zhou, X.: EdgeVision: Towards Collaborative Video Analytics on Distributed Edges for Performance Maximization. IEEE Transactions on Multimedia **26**, 9083–9094 (2024). https://doi.org/10.1109/TMM.2024.3385678

[19] Dai, X., Zhang, Z., Yang, P., Xu, Y., Liu, X., Lui, J.C.S.: AxiomVision: Accuracy-Guaranteed Adaptive Visual Model Selection for Perspective-Aware Video Analytics. In Proceedings of the 32nd ACM International Conference on Multimedia, pp. 7229–7238 (2024). https://doi.org/10.1145/3664647.3681269

[20] Liu, Z., Wang, Y., Zhao, Y., Qiu, C., Zhang, C., Wang, X., Dong, M.: Enabling Real-Time Video Detection With Adaptive and Distributed Scheduling in Mobile Edge Computing. IEEE Transactions on Mobile Computing **24**(12), 12784–12801 (2025). https://doi.org/10.1109/TMC.2025.3588142

[21] Kaul, S., Yates, R., Gruteser, M.: Real-time status: How often should one update? In Proceedings of INFOCOM'12, pp. 2731–2735 (2012). https://doi.org/10.1109/INFCOM.2012.6195689

[22] Chiariotti, F., Holm, J., Kalør, A.E., Soret, B., Jensen, S.K., Pedersen, T.B., Popovski, P.: Query Age of Information: Freshness in Pull-Based Communication. IEEE Transactions on Communications **70**(3), 1606–1622 (2022). https://doi.org/10.1109/TCOMM.2022.3141786

[23] Maatouk, A., Assaad, M., Ephremides, A.: The Age of Incorrect Information: An Enabler of Semantics-Empowered Communication. IEEE Transactions on Wireless Communications **22**(4), 2621–2635 (2023). https://doi.org/10.1109/TWC.2022.3213227

[24] Meesena, W., Nikunram, C., Turner, S.J., Supittayapornpong, S.: Minimizing Age of Processed Information Over Unreliable Wireless Network Channels. IEEE Transactions on Mobile Computing **24**(5), 3567–3578 (2025). https://doi.org/10.1109/TMC.2024.3520913

[25] Xie, S., Ma, S., Ding, M., Shi, Y., Tang, M., Wu, Y.: Robust Information Bottleneck for Task-Oriented Communication With Digital Modulation. IEEE Journal on Selected Areas in Communications **41**(8), 2577–2591 (2023). https://doi.org/10.1109/JSAC.2023.3288252

[26] Ma, S., Qiao, W., Wu, Y., Li, H., Shi, G., Gao, D., Shi, Y., Li, S., Al-Dhahir, N.: Task-Oriented Explainable Semantic Communications. IEEE Transactions on Wireless Communications **22**(12), 9248–9262 (2023). https://doi.org/10.1109/TWC.2023.3269444

[27] Wu, W., Yang, Y., Deng, Y., Aghvami, A.-H.: Goal-Oriented Semantic Communications for Robotic Waypoint Transmission: The Value and Age of Information Approach. IEEE Transactions on Wireless Communications **23**(12), 18903–18915 (2024). https://doi.org/10.1109/TWC.2024.3424493

[28] Wang, S., Bi, S., Zhang, Y.-J.A.: Edge Video Analytics With Adaptive Information Gathering: A Deep Reinforcement Learning Approach. IEEE Transactions on Wireless Communications **22**(9), 5800–5813 (2023). https://doi.org/10.1109/TWC.2023.3237202

[29] Wang, S., Yang, J., Bi, S.: Adaptive Video Streaming in Multi-Tier Computing Networks: Joint Edge Transcoding and Client Enhancement. IEEE Transactions on Mobile Computing **23**(4), 2657–2670 (2024). https://doi.org/10.1109/TMC.2023.3263046

[30] Dai, P., Chao, Y., Wu, X., Liu, K., Guo, S.: Context-aware offloading for edge-assisted on-device video analytics through online learning approach. IEEE Transactions on Mobile Computing **23**(12), 12761–12777 (2024). https://doi.org/10.1109/TMC.2024.3418608

[31] Kong, Y., Yang, P., Cheng, Y.: Edge-Assisted On-Device Model Update for Video Analytics in Adverse Environments. In Proceedings of the 31st ACM International Conference on Multimedia, pp. 9051–9060 (2023). https://doi.org/10.1145/3581783.3612585

[32] Nan, Y., Jiang, S., Li, M.: Large-Scale Video Analytics with Cloud–Edge Collaborative Continuous Learning. ACM Transactions on Sensor Networks **20**(1), 14–11423 (2024). https://doi.org/10.1145/3624478

[33] Wang, Z., Zhang, R., Zhang, S., Cheng, W., Wang, W., Cui, Y.: Edge-Assisted Adaptive Configuration for Serverless-Based Video Analytics. IEEE Transactions on Networking **33**(3), 1144–1159 (2025). https://doi.org/10.1109/TON.2024.3523956

[34] Ayyoubzadeh, S.M., Liu, W., Kezele, I., Yu, Y., Wu, X., Wang, Y., Jin, T.: Test-time adaptation for optical flow estimation using motion vectors. IEEE Transactions on Image Processing **32**, 4977–4988 (2023). https://doi.org/10.1109/TIP.2023.3309108

[35] Yan, Y., Zhang, S., Jin, Y., Cheng, F., Qian, Z., Lu, S.: Spatial and temporal detection with attention for real-time video analytics at edges. IEEE Transactions on Mobile Computing **23**(10), 9254–9270 (2024). https://doi.org/10.1109/TMC.2024.3361016

[36] Shi, X., Zhang, S., Wu, J., Chen, N., Cheng, K., Liang, Y., Lu, S.: Adapyramid: Adaptive pyramid for accelerating high-resolution object detection on edge devices. IEEE Transactions on Mobile Computing **23**(8), 8208–8224 (2023). https://doi.org/10.1109/TMC.2023.3343448

[37] Li, X., Zhang, S., Huang, Y., Ma, X., Wang, Z., Luo, H.: Towards timely video analytics services at the network edge. IEEE Transactions on Mobile Computing **23**(11), 10443–10459 (2024). https://doi.org/10.1109/TMC.2024.3376769

[38] Dendorfer, P., Osep, A., Milan, A., Schindler, K., Cremers, D., Reid, I., Roth, S., Leal-Taixé, L.: MOTChallenge: A Benchmark for Single-Camera Multiple Target Tracking. International Journal of Computer Vision **129**(4), 845–881 (2021). https://doi.org/10.1007/s11263-020-01393-0

[39] Varga, L.A., Kiefer, B., Messmer, M., Zell, A.: SeaDronesSee: A Maritime Benchmark for Detecting Humans in Open Water. In Proceedings of 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 2260–2270 (2022). https://doi.org/10.1109/WACV51458.2022.00374

[40] Cao, Y., He, Z., Wang, L., Wang, W., Yuan, Y., Zhang, D., Zhang, J., Zhu, P., Van Gool, L., Han, J., et al.: VisDrone-DET2021: The Vision Meets Drone Object Detection Challenge Results. In Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), pp. 2847–2854. https://doi.org/10.1109/ICCVW54120.2021.00319

[41] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., Ding, G.: YOLOv10: Real-Time End-to-End Object Detection. In Proceedings of 38th Conference on Neural Information Processing Systems (NeurIPS 2024), pp. 107984–108011 (2024). https://doi.org/10.52202/079017-3429