1 Vision and Research Summary

We need a robust **Automated Semantic Networking** framework to guide the design of **Data Sensing, Scheduling and Routing**. Current communication networks encounter the following obstables in supporting emerging applications such as the autonomous driving and AR/VR: (1) Processing and analyzing the **large amount of multimodal data** is time-consuming and challenging; (2) Handcrafting the network operating parameters for applications with **heterougenous QoS requirement in various environment** is inefficient.

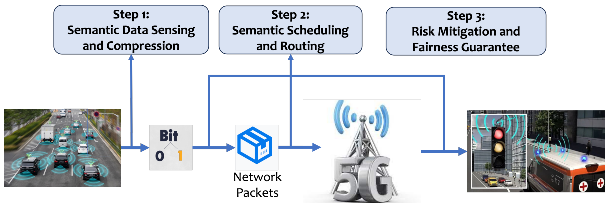

Figure 1: My research map on automated semantic networking: data-driven optimization and risk mitigation.

My research utilizes and develops new tools in the field of **stochastic optimization, information theory and generative models** to design efficient strategies for: 1. Controlled semantic information sensing and data retrieval; 2. Network control algorithms considering packet level semantic importance; 3. Deployment issues including risk mitigation and fairness guarantees. Fig. 1 illustrates the structure of my research. Fig. 2 summarizes my past results and future directions. Details of my research plan can be found in Section 2–4.

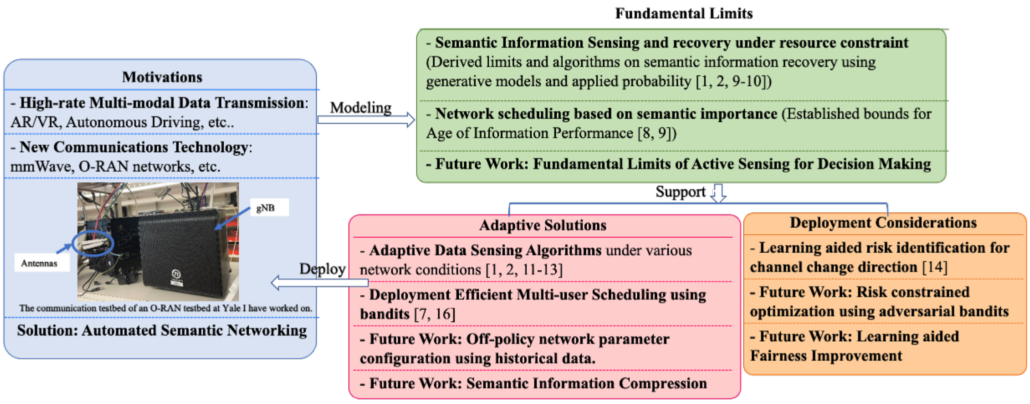

Figure 2: My research outlook and past results on semantic networking: from data sensing and scheduling to risk mitigation.

2 Semantic Sensing and Data Recovery

Extracting and transmitting semantic information that is useful for task completion is the key to support various applications such as AR/VR under limited communication resources. To achieve this, it is important to first sample the most important data, and then develop efficient compression algorithms before sending them out to the communication networks. To date, I have made the following research progress in semantic data sampling and retrieval:

(1). **Online sampling for freshness optimization under resource constraint**: By using the Age of Information (AoI) as a metric to quantify data freshness from the perspective of the receiver, we propose an **online** sampling and flow control algorithm that can minimize the average AoI at the receiver [1, 12]. Our algorithm can operate without knowing the upper and lower bound of delay expectation, while still preserves the almost sure convergence property in the original Robbins-Monro algorithm. Moreover, by using the Le Cam's two point method from non-parametric statistics, we then show our algorithm is **order optimal** under the worst case delay distribution. This algorithm can be generalized to the sampling and flow control problem in a wide range of time-varying processes such as the Wiener and Poisson process [2, 13].

(2). **High-dimention semantic data recovery using generative models:** We developed a novel encoder/decoder training methodology using the compatibility of joint distribution [9]. The proposed model can encode high dimension images into a low dimension latent space, and is also capable of encoding and generating new data. By leveraging a pretrained diffusion generative model, we propose a new method to retrieve low-resolution, blurred images using techniques from gradient estimation [10]. Our method does not need task specific model training of fine-tuning on the generative model.

In the future, I want to pursue research in the following problems:

(1). **Task-oriented semantic active sensing:** A more comprehensive assessment of the quality of each sampled data packet should emphasise its importance to task completion, such as the improvement in the planning efficiency in autonomous driving. This impact cannot be meausred directly using conventional metrics such as the MSE. I plan to quantify the importance of each sample to task completion by computing the uncerntainty reduction or information gain in each decision making task, then propose active learning methods to sense the most informative data.

(2). **Semantic neural data compression:** The sampled data can also be divided into multiple patches. We aim at using learning algorithm to find the most informative patch, and propose solutions based on generative modeling to compress and recover the informative patches. My past research on semantic data recovery and cyclic generation will serve foundations for this thrust.

3 Network Scheduling based on Semantic Importance

The semantic importance of each packet should also be considered as an index in designing network scheduling algorithms. By using data freshness as an semantic measure, I have made the following progress:

(1). **Multi-user freshness oriented scheduling:** My work [3, 4] propose the first low-complexity joint data sampling and cross-layer transmission algorithm that is **AoI optimum** in networks with massive number of access nodes. By further taking the time-varying user and task requests into account, we develop in [5, 6] cache updating strategies so that users can fetch the freshest copies of time-varying files from the local cache. To catch the freshest sporadic updates (e.g., anomaly behaviors), by using the Age of Synchronization (AoS) as a freshness metric, we propose an AoS minimum uni-cast algorithm in multi-user networks and show that sources with smaller update frequency should be guaranteed transmission priority [7, 8].

(2). **Deployment efficient learning for scheduling:** Online learning algorithm can find the optimum scheduling strategy under unknown network statistics. We first improve the explore-exploit efficiency in online queueing systems using information directed sampling [16]. Then we improved the deployment efficiency of online link rate adaptation algorithms using batched bandits [15].

Based on the initial success on freshness oriented network, I plan to improve and extend the current semantic scheduling framework as follows:

Improving the generalization performance of scheduling algorithms in new carriers and networks. During the initial parameter selection phase for networks operating in new frequency bands, our plan is to discern the causal relationship between network conditions and network parameters by leveraging historical data. This will enable us to tune network parameters from a good starting point. To further improve the network paramter configurations in operation phase, we plan to improve the accuracy of the QoS evaluation using both online data collected from the new carrier and historical data. My past work on causal inference [17] will serve as a research foundation in this thrust.

4 Risk Mitigation and Fairness Guarantees

Networking algorithms need to be robust before they are deployed to the real world systems. Moreover, networking algorithms need to be fair while guaranteeing a high operating efficiency and utility. In the future, I plan to mitigating the operation risk and improve the faireness as follows:

(1). **Provable risk constrained multi-objective optimization:** Besides optimizing the average network system performance over a long period, I believe it is crucial to develop algorithms whose risk (e.g., channel outage probability) will not exceed a certain threshold provably. To develop such risk constrained optimization algorithms, combined with the automated network parameter configuration framework, I intend to establish the lower bound on performance regret under the fact that a certain amount of risk events may happen, and develop risk aware adaptive learning algorithms for performance improvements.

(2). **Learning for fairness guarentees:** It is well known that an elephant flow may take up the majority of the bandwidth when competing with a mice flow. Similarly, we observe that a throughput optimal flow may take up the majority of the bandwidth when sharing a First-Come-First-Serve router with a time-sensitive flow [11]. To prevent the unfair phenomena, we aim at design incentive mechanism for routers ad controllers in the O-RAN networks by using tools for multi-agent reinforcement learning.

5 Looking Ahead

The long-term goal of my research is to support diverse autonomous applications using B5G/6G networks systems using information theory, statistical inference, learning, and signal processing. Learning and Collaborating with peers are the essential way I approach my research goal. I have a variety of academic collaborators, including junior and senior colleagues at Yale University, Auburn University and industry partners such as Microsoft. These collaborations have strengthened the work itself and helped me grow and develop my research vision.

Bibliography[1]

[1] **H. Tang**, Y. Chen, J. Wang, P. Yang and L. Tassiulas, "Age Optimal Sampling Under Unknown Delay Statistics", *IEEE Transactions on Information Theory*, Feb. 2023, pp. 1295-1314.

[2] **H. Tang**, Y. Sun and L. Tassiulas, "Sampling of the wiener process for remote estimation over a channel with unknown delay statistics", accepted *IEEE/ACM Transactions on Networking*.

[3] **H. Tang**, J. Wang, L. Song and J. Song, "Minimizing Age of Information With Power Constraints: Multi-User Opportunistic Scheduling in Multi-State Time-Varying Channels," in *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 5, pp. 854-868, May 2020.

[4] **H. Tang**, J. Wang, L. Song and J. Song, "Scheduling to Minimize Age of Information in Multi-State Time-Varying Networks with Power Constraints," *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2019, pp. 1198-1205.

[5] **H. Tang**, P. Ciblat, J. Wang, M. Wigger and R. D. Yates, "Cache Updating Strategy Minimizing the Age of Information with Time-Varying Files' Popularities," *2020 IEEE Information Theory Workshop (ITW)*, 2021, pp. 1-5.

[6] **H. Tang**, P. Ciblat, J. Wang, M. Wigger and R. Yates, "Age of Information Aware Cache Updating with File- and Age-Dependent Update Durations," *2020 18th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, 2020, pp. 1-6.

[7] **H. Tang**, J. Wang, Z. Tang and J. Song, "Scheduling to Minimize Age of Synchronization in Wireless Broadcast Networks With Random Updates," in *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 4023-4037, June 2020.

[8] **H. Tang**, J. Wang, Z. Tang and J. Song, "Scheduling to Minimize Age of Synchronization in Wireless Broadcast Networks with Random Updates," *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 1027-1031.

[9] C. Liu, **H. Tang**, T. Qin, J. Wang and T. Y. Liu, "On the Generative Utility of Cyclic Conditionals". *Advances in Neural Information Processing Systems (NeuIPS)*, 2021, 34: 30242-30256.

[10] **H. Tang**, X. Xie, A. Feng, H. Wang, C. Zhang and Y. Bai, "Solving Inverse Problems with Diffusion Generative Models: a Policy Gradient Viewpoint," in *NeurIPS Workshop on Deep Learning and Diffusion Models*

[11] Y. Tan*, **H. Tang**, J. Wang and J. Song, "Early Drop: A Packet-Dropping Incentive Rate Control Mechanism to Keep Data Fresh under Heterogeneous QoS Requirements," *2021 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2021, pp. 1-6.

[12] **H. Tang**, Y. Chen, J. Sun, J. Wang and J. Song, "Sending Timely Status Updates through Channel with Random Delay via Online Learning," *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications*, 2022, pp. 1819-1827.

[13] **H. Tang**, Y. Sun and L. Tassiulas, "Sampling of the wiener process for remote estimation over a channel with unknown delay statistics". In *Proceedings of the Twenty-Third International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (MobiHoc '22)*. Association for Computing Machinery, New York, NY, USA, 51–60.

[14] **H. Tang**, X. Hou, J. Wang and J. Song, "Joint Link Rate Selection and Channel State Change Detection in Block-Fading Channels," *2021 IEEE Global Communications Conference (GLOBECOM)*, 2021, pp. 1-6.

[15] Y. Chen*, **H. Tang**, J. Wang and J. Song, "Adaptive Link Rate Selection for Throughput Maximization with Batched Thompson Sampling," *2022 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2022, pp. 1-6.

[16] X. Hou*, **H. Tang**, J. Wang and J. Song, "Information Directed Learning Algorithm for Minimizing Queue Length Regret," *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2022, pp. 1-5.

[17] C. Liu, X. Sun, J. Wang, T. Li, **H. Tang**, T. Qin, T. Y. Liu, "Learning causal semantic representation for out-of-distribution prediction", *Advances in Neural Information Processing Systems (NeuIPS)*, 2021, 34: 6155-6170.

[1]* denotes undergraduates and junior graduate students working on projects with me.