Research Summary

Xinyu Wu

January 9, 2021

1 Finite-Buffer Communication System Analysis

Time: Sep.2019 - Present (Ongoing)

Advisor: Prof. Eytan Modiano in MIT

Introduction: Finite buffer systems appear ubiquitously due to the storage resource constraints. For example, buffer size is limited to be small in networks on-chip, due to the area and power restriction, as well as in computer interconnection networks and spacecraft networks. Previous works focus primarily on infinite-buffer systems and propose a systematic framework for analyzing their throughput, stability, etc. The finite-buffer systems, however, are not thoroughly and systematically explored. Nevertheless, the following toy example in Fig. 1 unveils that finite-buffer systems may have fundamental difference with infinite-buffer systems, specifically reducing the throughput. In view of this simple observation, in this series of works we are motivated to build a generalized framework to analyze the effect of finite buffer on the communication network systems in terms of the stability, throughput, fairness etc.

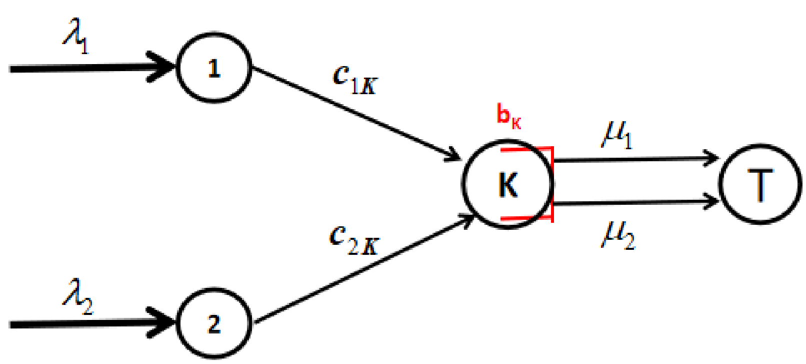


Figure 1: Example of Finite Buffer Reducing Throughput. The buffer of node K accepts both commodity 1 and 2, with buffer size b_K. Node T is the destination for both commodities. Assume that $c_{1K} > \mu_1$, $c_{2K} > \mu_2$, $q_T(t) \equiv 0$, and the transmission policy is that for each link, whenever there exists commodity backlog in the source node of the link, then it can be transmitted with rate equal to the capacity value of the link, otherwise any commodity cannot be transmitted. We can observe that if $b_K = \infty$, then the maximum throughput for commodity ℓ is μ_ℓ. However if $b_K < \infty$ and $c_{1K}/\mu_1 > c_{2K}/\mu_2$, when commodity 1 is overloaded, i.e. $\lambda_1 > \mu_1$, then the maximum throughput for commodity 2 is reduced to $(\mu_1/c_{1K})\mu_2 < \mu_2$.

Specifically, our ongoing contribution is summarized as follows:

- We proposed an ordinary differential equation (ODE) model to characterize the dynamics of queue-based model in the finite-buffer system, and proved the sufficient condition of the transmission policy that can stabilize the networks when the flow arrival vector is strictly inside the admissible rate region, for both single and multi-commodity cases.

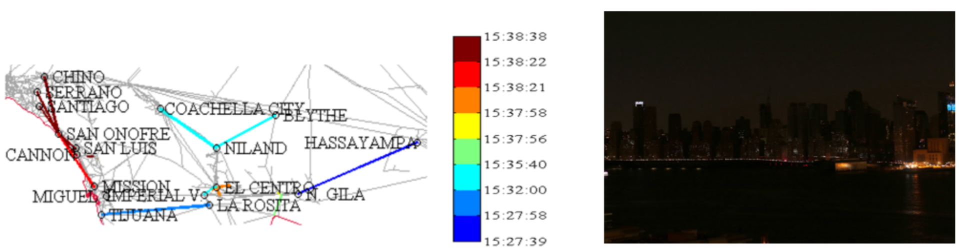
- We figured out the finiteness of buffer size may induce throughput reduction under the overloaded multi-commodity system, and we formulated an optimization framework to capture the amount of throughput reduction and even conduct the capacity allocation such that the throughput can be maximized.
- We proved that the backpressure policy can achieve the most balanced overloading rate vector under the finite-buffer system with given capacity values. We then clarified the set of conditions of link capacity values to achieve the most balanced overloading rate, and showed that a greedy-based mechanism in terms of shortest path can obtain the optimal capacity allocation to achieve these conditions.

2 Failure Cascade Prediction and Analysis in Power Systems

Time: Nov.2018 - Sep.2019

Advisor: Prof. Eytan Modiano in MIT

Introduction: Modern power systems frequently experience unpredictable component failures which are caused by falling tree branches, storms, aged devices, wrong protective actions, etc. Past events have shown that large scale blackouts are usually accompanied by a rapid propagation of failures among many system components, where real examples are shown in Fig. 2. This "rolling snowball" phenomenon is referred to as a *failure cascade*. Modeling and predicting failure cascades is very difficult because a large scale power grid can have hundreds of thousands of components whose dynamic interactions are nonlinear and whose parameters are varying with time. In this work, we target at applying the influence model to characterize, predict and analyze the failure cascade dynamics in power systems. The influence model is a Markovian-like model that captures the underlying influence of all the network components on each individual component.



San Diego Power Blackout on Sept. 8, 2011

A 500 kV line tripped off by a mistake.
Wrongly cut off 2 generators in Mexico.
Finally separated into three islands which collapsed afterwards.

Manhattan City Blackout on Jul. 13, 2019

A transformer fired at West 64th Street and West End Avenue

Figure 2: Examples of Large-Scale Failure Cascade

Specifically, our contribution is summarized as follows:

- We characterized the failure cascade process in power systems by the influence model (IM), and proposed a hybrid learning scheme to train IM. First we applied a Monte-Carlo approach to quickly acquire the pairwise influence between any two transmission links; Then we formulated a quadratic programming to obtain the weights of each pairwise influence; Finally we proposed an adaptive threshold estimation and selection scheme to better predict cascade processes.
- We tested our prediction mechanism based on the 1354-bus, 2383-bus, and 3012-bus systems provided by IEEE in Matlab MATPOWER Toolbox, under DC and AC flow. Our mechanism could predict the final state of each link within 10% error rate, final failure size within 7% error rate, and failure time within 1 time unit generally, with at most 3.4% of all possible initial link contingencies as training samples and two magnitudes faster in prediction than the flow calculation method.

- We proposed an efficient way to identify the most critical initial contingency that can lead to large failure range based on the learned influence model.

Publications:

- Xinyu Wu, Dan Wu, Eytan Modiano, "Predicting Failure Cascades in Large Scale Power Systems via the Influence Model Framework", submitted to IEEE Transactions on Power Systems, under review.
- Xinyu Wu, Dan Wu, Eytan Modiano, "An Influence Model Approach to Failure Cascade Prediction in Large Scale Power Systems", the 2020 American Control Conference (ACC), Denver, USA, Jul. 1st-3rd, 2020.

3 Social Network De-Anonymization

Time: Mar.2017 - Aug.2018

Advisor: Prof. Luoyi Fu, Prof. Xinbing Wang in Shanghai Jiao Tong University

Introduction: The problem of social network de-anonymization is to identify users in an anonymized network by another cross-domain un-anonymized network based on auxiliary information such as topological similarity. Fig. 3 shows a toy example of two social networks, one un-anonymized (left) and one anonymized (right). The correlation of these two networks may benefit the identification of users in the anonymized networks, for example one user is the central node of one network that connects to a bunch of users tends to have large degree in another network; two users in one network belong to one community may share one community as well in another network. In this work, we target at the theoretical, algorithmical, and experimental analysis over the de-anonymization problem with overlapping community structure.

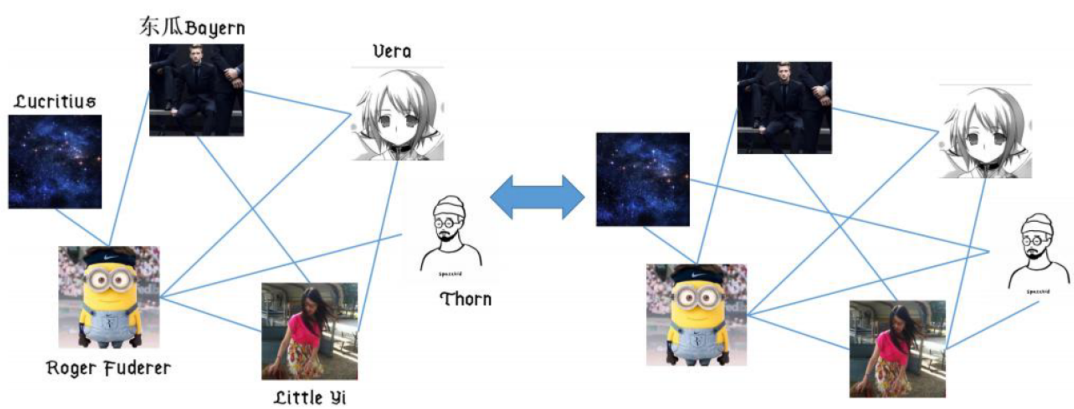


Figure 3: Toy Example of Social Network De-anonymization

Specifically, our contribution is summarized as follows:

- We formulated the cost function based on the Minimum Mean Square Error criterion, proved the NP-hardness of minimizing the cost function by reduction from 1-median problem, and derived an approximated alternative to this problem via the restriction by Sequence Inequality, with the approximation ratio at most 2.
- We showed that overlapping communities have a positive impact on the de-anonymization accuracy by matrix theory, and proposed the Convex-concave Based De-anonymization Algorithm (CBDA) to solve the alternative problem.

- We confirmed the effectiveness of CBDA under the Microsoft Academic Graph with 3176 nodes and 89 overlapping communities, which showed an average of 90% accuracy in correctly identifying users in un-anonymizee networks when communities overlap densely, with roughly 70% enhanced identification ratio compared with non-overlapping cases.

Publication:

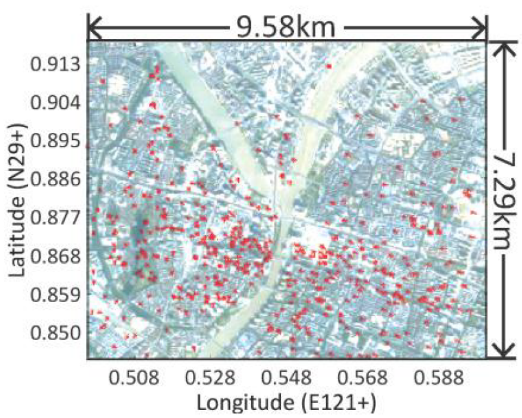
- Xinyu Wu, Zhongzhao Hu, Xinzhe Fu, Luoyi Fu, Xinbing Wang, Songwu Lu, "Social Network De-Anonymization with Overlapping Communities: Analysis, Algorithm, and Experiments", IEEE International Conference on Computer Communications (INFOCOM), Honolulu, USA, Apr. 15th-19th, 2018.

4 Large-Scale Wireless Fingerprinting Prediction and Localization

Time: Sep.2016 - Mar.2017

Advisor: Prof. Xiaohua Tian, Prof. Xinbing Wang in Shanghai Jiao Tong University

Introduction: Cellular network positioning is a mandatory requirement for localizing emergency callers, such as E911 in North America. Although smartphones are normally equipped with GPS modules, there are still a large number of users with cell phones only as basic devices, and GPS could be ineffective in urban canyon environments. To this end, the RF fingerprints based positioning mechanism is incorporated into LTE architecture by 3GPP, where the fingerprint denotes the wireless signal strength from different base stations. The obstacles in urban area disenable the full collection of geo-tagged RF fingerprints. This paper aims to utilize the subspace identification approach for large-scale RF fingerprints prediction over the whole targeted region given the limited available samples, and then apply the predicted fingerprints for outdoor positioning.



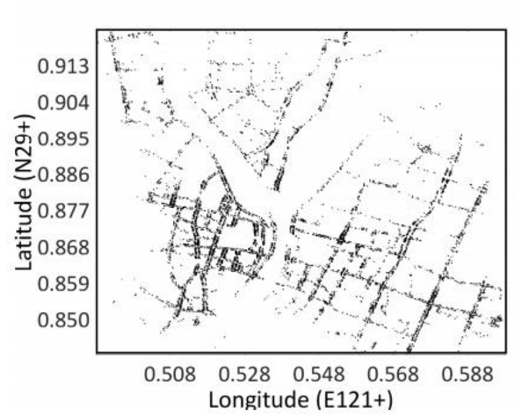


Figure 4: Base Station Distribution (left) and Available Fingerprint Sample Distribution (right, principally on the main roads) in Ningbo, China. We aim to predict the fingerprints of the blank area in the right subfigure given the collected samples on the main roads.

Specifically, our contribution is summarized as follows:

- We modeled the fingerprint prediction as a matrix completion problem, and proposed a Stiefel-manifold based algorithm for prediction based on Singular Value Decomposition and QR Decomposition.

- We further designed a sliding-window mechanism to overcome the sparsity of fingerprints that solely gathered on main roads, by means of firstly applying our method in small windows around main roads, and gradually percolating to the whole region with larger windows as more fingerprints had been predicted.
- We reconstructed the fingerprints in a 69.8km^2 region in Ningbo, China, shown in Fig. 4, and showed that 71% and 98% users can respectively be localized within an error of 100m and 300m, triumphing over the accuracy of Cell-ID and Gaussian Mixture Model approaches, and achieving E911's requirements: "within 100m for 67% and 300m for 90%."

Publications:

- Xiaohua Tian, Xinyu Wu, Hao Li, Xinbing Wang, "RF Fingerprints Prediction for Cellular Network Positioning: A Subspace Identification Approach", IEEE Transactions on Mobile Computing, vol. 19, no. 2, pp. 450-465, 2019.
- Xinyu Wu, Xiaohua Tian, Xinbing Wang, "Large-scale Wireless Fingerprints Prediction for Cellular Network Positioning", IEEE International Conference on Computer Communications (INFOCOM), Honolulu, USA, Apr. 15th-19th, 2018.