RESEARCH STATEMENT: JIAN TAN

jt2047@columbia.edu

My research interests range from the areas of electrical engineering and computer science to stochastic operations research and applied probability, with special emphasis on large scale information networks and service systems. I focus on understanding the fundamental principles that govern the operations of these systems and using these underlying laws to improve the systems design. More specifically, my thesis work focuses on the design of scheduling algorithms for communication/service networks and the interplay of protocol designs and origins of heavy-tailed distributions therein; this work can be organized in the following three general topics: 1) origins of heavy-tailed (e.g., power law) distributions; 2) redesign of network protocols to prevent the cause (or reduce the effects) of heavy tails and 3) design of new adaptive (self-organizing) scheduling algorithms. Recently (September 2009), in the Multi-University-Research-Initiative (MURI) project meeting that involves multiple major networking theory groups, my work on retransmissions causing heavy-tails drew lots of attention from a number of well-recognized researchers in the theoretical networking community. Outside of my thesis, I also did research on sensor networks and information theoretic security that are briefly described at the end of this research summary.

Origins of heavy-tailed distributions Since the early works of Pareto in 1897 and later of Zipf, heavy tails have been repeatedly observed for over a hundred years. Heavy-tailed distributions, in particular power laws, have been found in a variety of biological, technological and socioeconomic areas. Similarly to the fact that the wide appearance of Gaussian/Normal distributions can be attributed to the generality of the central limit theorem, in this research, I have investigated new universal laws that under very general conditions invariably result in heavy tails.

1. *Proportional growth and modulated branching process.* Physical objects that grow proportional to their size, e.g., computer files, Web sites, computer networks, city populations, sizes of companies and living organisms, wealth, etc, almost exclusively give rise to power law distributions. After carefully examining the proportional growth situations, we discover that most of them are characterized by the following three features. First, the objects of interest evolve due to the replication of their many independent components. Secondly, the rate of replication of the many components is often controlled by exogenous parameters causing periods of high growth and contraction, e.g., economic growth and recession, etc. Thirdly, the sizes of these objects often either have reflective lower boundaries, e.g., cities do not fall bellow a certain size, low income individuals are subsidized by the government, companies are protected by bankruptcy laws, etc; or have porous/absorbing lower boundaries, e.g., cities may degenerate, bankruptcy protections may fail and companies can be liquidated. Hence, we study a new class of models termed modulated branching processes with reflective or absorbing barriers. We show that these classes of models, under very general polynomial Gärtner-Ellis conditions, result almost invariably in power laws [C1,J2] (as listed in my CV). Informally, the interpretation of our main

results suggests that alternating periods of growth and reduction, e.g., economic expansions and recessions, are primarily responsible for the appearance of power law distributions. After completing the work, it was pointed to us by Professor David Aldous that our results are related to one of his proposed open problems; see http://www.stat.berkeley.edu/users/aldous/Research/OP/brw.html.

2. *Heavy-tailed delays induced by retransmissions.* Retransmissions represent one of the most fundamental approaches in communication networks that guarantee data delivery in the presence of channel failures. We have recognized in a sequence of papers [C3,C4,C7] that retransmissions may introduce heavy-tailed delays and possibly instabilities on all networking layers. This is the first body of work that attributes heavy-tails/long-range dependence to protocol designs. Furthermore, we have rigorously uncovered the detailed structure of the class of heavy-tailed distributions induced by retransmissions [C8,J3]. More precisely, we studied how the functional dependence between the data size variability and the channel dynamics impacts the distributions of the number of retransmissions N and the total transmission time T. In this functional space, we discover functional criticality points that naturally separate the asymptotic behavior of the distributions of N and T into the classes of very heavy (including power law distributions), moderately heavy (Weibull distributions) and nearly exponential distributions. Interestingly, the functional behavior of $P[N > n]$ can change even inside these classes. This work provides a new explanation of the widely observed heavy-tailed delays in communication networks since channel failures, variable data sizes and retransmissions are inherent part of modern network architectures.

Redesign of network protocols to prevent the cause of heavy tails Based on our new theoretical results, we show how the retransmission-based protocols may result in power law delays and instabilities on all networking layers, even if the transmitted data units have very concentrated distributions, e.g., exponential/Gaussian. These findings suggest that a careful reexamination and redesign of retransmission-based protocols in current and future networks is needed. Our theoretical predictions are in agreement with the reported empirical measurements, which have shown that the utilization of the 802.11 protocol is only 40%, basically due to retransmissions.

1. *Data Link Layer.* In case of an error, the packet is resent by Automatic Repeat reQuest (ARQ) protocol. We studied a situation when packets of variable sizes are sent over a single channel modeled by an on-off process [C3]. We showed that power law delays might arise if the hazard function of the packet size variability and channel available periods are proportional, irrespective of how light or heavy the distribution of the packet sizes may be.

2. *Media Access Control (MAC) Layer.* A similar phenomenon could arise for ALOHA-based MAC protocols, though the system is more complicated than the previous single user model. This work [C4], which won the Best Student Paper Award at ITC'20, discovered a new phenomenon that a basic finite population ALOHA model

with variable size (exponential) packets is characterized by power law transmission delays, possibly even resulting in zero throughput. Furthermore, we identified a new stability condition for contention based systems that is derived from the tail behavior of the packet and backoff distributions, contrary to the traditional results that are based exclusively on mean values. This power law effect might be diminished, or perhaps eliminated, by reducing the variability of packets. However, we also show that even a slotted (synchronized) ALOHA with packets of constant size can exhibit power law delays when the number of active users is random.

3. *Transport Layer.* We discovered that the end-to-end acknowledgements that have been widely used in the transport layer could also result in heavy-tails [C7], which may even lead to zero throughput. Thus, we argue that engineering data transport over the long routes using end-to-end acknowledgements requires a more careful design.

In order to alleviate the power law effect, we propose a dynamic packet fragmentation technique that can be easily implemented while at the same time reducing the total transmission time [C9,J6]. Also, we are currently investigating possible redesigns of ALOHA type protocols and end-to-end acknowledgement schemes.

Scheduling algorithms for the heavy-tailed environment Scheduling strategies play an important role in modern communication networks since one can dramatically improve the system performance by carefully assigning the service order to the requests. In this work, we make use of the observed heavy-tailed traffic characteristics and design scheduling algorithms that fit these constraints. We proposed a new adaptive, low-complexity and scalable scheduling algorithm [C5,J1], termed comparison scheduling, that can be shown in a certain sense to be close to the optimal. This algorithm is based on relative comparison of a newly arriving job to the sizes of the previous m arrivals. The adaptive and scalable nature of this comparison mechanism enables the scheduling algorithm to work well even when the traffic characteristics are nonstationary, highly correlated (long range dependent) and bursty (e.g., batch arrivals, etc). Furthermore, by introducing a new notion of conditional limits, we show that the sojourn time distribution of SRPT exhibits superior asymptotic performance as compared to PS and FBPS scheduling disciplines [C2]. This intuitively appealing result improved the previous understanding that the delay distributions in SRPT, PS and FBPS behave asymptotically the same in the presence of heavy-tails.

Sensor networks During my internship at the mathematical and algorithmic sciences research center of Lucent Bell Labs, I investigated a new class of anchoritic sensor networks [C6] that solves the localization problem by exploring the statistical properties of noise, which differs from the usual belief that noise is always detrimental to the system and should be preferably filtered away. This work won the Best Paper Award in DCOSS'07.

Information theoretic security In recent years, the famous wiretap channel has been revisited and information theoretic secrecy has become an active area of research in this setting. We design a wireless communication system that achieves constant bit rate data transmission over a block fading channel, securely from an eavesdropper that listens to the transmitter over another independent block fading channel. In the classical wiretap setting, it is well known that information theoretic secrecy at a constant bit rate is not possible. The outages occur at times when the eavesdropper channel has favorable conditions over the main channel. In our system, however, we exploit the times at which the main channel is favorable over the eavesdropper channel for us to be able to transmit some random secret key bits along with the data bits. These key bits are stored in a separate key buffer at the transmitter as well as the receiver, and are utilized to secure data bits, whenever the channel conditions favor the eavesdropper. We develop the key buffer and power control mechanisms to achieve maximum secure constant bit rate, with which the outage probability can be made arbitrarily close to 0. This work can be viewed as a first step in providing a framework that combines both information theory and queueing analysis for the study of information theoretic security.

Future work My future work interests will also be composed of both theoretical and practical research elements. Specifically, due to my multi-disciplinary background in operations research and communication networks, I am interested in developing innovative ways to design, manage and enhance service systems (e.g., cloud computing), and especially in discovering fundamental principles and developing scientific/engineering innovations in that setting. Furthermore, I am excited to venture into other applied areas, e.g., large scale data management and computing systems, stochastic network optimization, machine learning with application to bioinformatics.