Preparing for the Worst: Applications of Stochastic Orders in Network Algorithmics

Paul Tune and Jun (Jim) Xu*

Abstract

Performing operations on the traffic of high-speed networks face the typical problem of volume, velocity and variety. Network algorithmics is a field developed to deal with these problems. Of importance to network algorithmics is the inputs/workloads into the solutions developed by the field are highly variable, and so is often modeled as a stochastic process. Moreover, a chief cornerstone of network algorithmics are randomized algorithms, which tradeoff a resource (for example, space) for a small, allowable error. Thus, it is crucial that one has to derive probabilistic guarantees on the performance of a solution with a worst-case workload under a performance metric (for instance, estimation error). This is usually achieved via probability bounds such as the famous Chernoff bound, but deriving such bounds for solutions from network algorithmics is often challenging.

The derivation of these bounds can be dramatically simplified by exploiting a stochastic order, which is a notion of when one random variable is "larger" than another. In this paper, we survey some stochastic orders and show how they can be used to provide simple proofs of these bounds in the design of randomized algorithms in network algorithmics.

Index terms— Log-concave distributions, majorization, network algorithmics, stochastic ordering, supermodularity.

1 Introduction

Over the past two decades, the field of *network algorithmics* [31] has emerged as a rich area of research. The types of problems that network algorithmics address include, but are not limited to, packet classification, queue management, traffic shaping, switching and routing, network measurement, and data streaming analysis. Many of such solutions have been successfully deployed in commercial networks, including solutions for Internet routers, security apparatus, and measurement devices.

In general, network operators would like solutions that are robust under a wide variety of, *often unforeseen*, operating conditions. However, the performance of a network appliance design or the accuracy of an estimation method is often dependent on workload uncertainties that are beyond the control of the operator. Unfortunately, applicable mathematics for the rigorous analysis of the worst-case stochastic behaviors of network algorithmics solutions under arbitrary workloads is largely lacking.

Understanding how a solution would behave in the worst-case, not just in the typical case, is important for two reasons. First, with suitable mathematics to characterize worst-case workloads, we can design solutions that will work well under any operating conditions, including those in which an adversary is trying to break the system, or under unexpected changes in the usage pattern. Second, more often than not, we have, surprisingly, found in our past efforts that delivering solutions that can guarantee high performance under the worst-case conditions cost only slightly more than designs that don't. However, coming up with such solutions hinges on our ability to understand the characteristics of the worst-case scenarios so that we can design around them.

Large deviation theory [8] on $\mathbb{R}$ is concerned with the probability that the sum of some random variables $S := \sum_{i=1}^{n} X_i$ will exceed a given threshold x, which in our contexts may correspond to processing or network capacity, resource constraint, or tolerable error bound. In worst-case large deviation problems, these random variables $X_1, X_2, \cdots, X_n$ are the functions of some large parameter vector $\mathbf{q} := [q_1, q_2, \cdots, q_m]$, and we wish to find the worst-case probability tail bounds of $S(\mathbf{q}) := \sum_{i=1}^{n} X_i(\mathbf{q})$ under all possible parameter settings $\mathbf{q} \in Q$. In other words, we would like to compute $\max_{\mathbf{q}\in Q} \Pr[S(\mathbf{q}) \geq x]$.

*The first author is with the School of Mathematical Sciences, The University of Adelaide, Australia, and the second is with the College of Computing, Georgia Institute of Technology (Email: paul.tune@adelaide.edu.au, jun.xu@cc.gatech.edu).

Establishing such worst-case tail bounds is important not only for network security applications where an adversary is in full or partial control of this vector, but also for non-security applications where we would like to know the worst-case system performance under all operating conditions or workloads. Obtaining such worst-case bounds is very difficult because the parameter space Q is typically gigantic. Although establishing tail bounds is often straightforward through Chernoff bounding techniques when a particular parameter setting $\mathbf{q}$ is given, such bounds typically cannot be expressed as a closed form function of $\mathbf{q}$, rendering conventional optimization techniques powerless for maximizing $\Pr[S(\mathbf{q}) \geq x]$. Without worst-case large deviation machineries that have been or are being developed, the only conceivable option would be to enumerate all parameter settings $\mathbf{q} \in Q$, but repeating tail bound analysis over the entire parameter space Q is usually computationally prohibitive.

Whether determining the worst case input or bounding the behavior of randomized algorithms, *stochastic orders* can help in understanding the worst-case behavior of proposed solutions in network algorithmics. The motivation of this paper is to provide an overview to researchers in network algorithmics on stochastic orders.

A stochastic order basically defines when a random variable is "larger" than another. The precise definition of "larger" than will depend on the order. There are many types of stochastic orders that can be defined on random variables. Two chief references are Müller and Stoyan [27], and Shaked and Shantikumar [29]. Of importance to us are majorization and Schur convexity, supermodular and convex orderings. We will also briefly mention about negatively associated random variables and how they relate to network algorithmics.

In this paper, we will summarize some useful classifications of distributions and stochastic orders to aid in the design of randomized algorithms in network algorithmics. In particular, we cover:

(i) majorization and Schur convexity,

(ii) negative association,

(iii) convex ordering and supermodularity, and

(iv) log-concave distributions.

We show how these can be applied to various applications in network algorithmics. Most of the results shown here can be found in literature, and so we endeavor to reference the original paper where the result originated from.

2 Background and Motivation

As mentioned in the introduction, the study of random processes are central in network algorithmics. Common mathematical tools to bound the worst-case behavior includes the Markov's inequality, Chebyshev's inequality, the Chernoff bound and Azuma-Hoeffding inequality [3, 15]. Overviews of these tools (and many others) can be found in [1, 25, 26].

Why do we care about finding strong, refined worst-case probability bounds? Or equivalently, can one just make do with coarse bound?

The question can only be answered in context: if the problem can tolerate a wide margin of "bad" events, then a refined worst-case probability bound is unnecessary. More often than not, however, due to the high-speed, large data volume environment, stronger bounds are a necessity.

Consider the issue of packets arriving out-of-order after traversing a switch. TCP-based applications are sensitive to this packet reordering. A TCP-based application will drop the set of packets it has received thus far, and issue a retransmission to the sender, causing wastage of bandwidth. Suppose as a designer of a switch, at worst, we know that packets will arrive out-of-order with probability $1/1000$. Then, roughly 1 out of a 1000 packets will be out-of-order. Suppose each packet is has size 100 bytes, then on a 100 Gbps link, it takes 8 ns to transmit a single packet. Then, we expect about 1 packet to be out-of-order every 8 μs. This may not be acceptable for some applications, so stronger bounds are required.

The lesson here is that the volume of data processed by solutions from network algorithmics are massive, and certainly large enough for rare events to be seen in a given sample set.

We summarize a few areas where strong bounds are required:

Distributed data streaming problems: Analyzing massive aggregate traffic streams through many high-speed nodes (or links) for detecting global events (*e.g.,* global elephants) that can spread themselves even and thin over these nodes (to avoid detection) or for estimating global statistics in a communication-efficient manner, known as *distributed data streaming*

(DDS), has emerged in recent years as an important research topic in network monitoring. We will show that worst-case large deviation issues arise naturally in the design and analysis of DDS algorithms, where we need to obtain a tight probability tail bound on its detection or estimation accuracy no matter how the global event or statistic is split into fragments of different sizes across these nodes. The design of such stochastically robust solutions for two important DDS problems remain open: estimation of the empirical entropy of and association rule mining over distributed data streams.

Large per-flow data structure problems: During a measurement period, tens of millions of application flows may pass through an Internet gateway. Many network algorithmics methods have been developed that make use of per-flow state information, *e.g.,* per-flow scheduling, stateful firewalls, and multi-packet deep packet inspection for intrusion and virus detection. However, it is very challenging to design massive flow-level data structures that are capable of providing extremely high throughput and low bounded access delay under arbitrary workloads, yet have relatively low costs. Randomized algorithmic solutions that can provide strong worst-case stochastic performance guarantees while using commodity memories require strong stochastic bounds on their behavior (see for instance [16, 33]).

Load-balanced switching problems: Load-balanced routers are an important class of switch architectures that provides scalability and throughput guarantees. However, the plain vanilla load-balanced routers do not preserve the order of packets even within a TCP or UDP flow, which could cause performance problems for many network applications [5, 6]. While forcing all packets within a TCP flow go down the same path (*e.g.,* determined by hashing on the flow identifier of each packet) inside a router will eliminate this packet reordering issue, we will show that the significant heterogeneity in flow rates and sizes seen in today's Internet will leave these routers severely load-imbalanced, defeating the very purpose for which they were designed.

In the rest of the paper, we discuss some useful stochastic orderings and properties of random variables that help with the development of tighter worst-case bounds.

3 Majorization and Schur Convexity

The concept of *majorization*, essentially a partial order on the distribution of random variables defined below, has seen widespread applications in various fields. The *de facto* reference for majorization is Marshall and Olkin's book [23]. We shall briefly discuss majorization here.

Let us define majorization and *Schur convex functions*. The following definitions can be found in [23].

Definition 1. *For any n-dimensional vectors $\mathbf{x}$ and $\mathbf{y}$, let $x_{[1]} \geq \cdots \geq x_{[n]}$ and $y_{[1]} \geq \cdots \geq y_{[n]}$ denote the components of $\mathbf{x}$ and $\mathbf{y}$ in non-increasing order respectively. We say $\mathbf{x}$ is majorized by $\mathbf{y}$, denoted by $\mathbf{x} \leq_M \mathbf{y}$ if $\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}$ for $k = 1, 2, \cdots, n-1$ and $\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]}$.*

Definition 2. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is Schur convex (Schur concave) if $\mathbf{x} \leq_M \mathbf{y}$ implies $f(\mathbf{x}) \leq f(\mathbf{y})$ ($f(\mathbf{x}) \geq f(\mathbf{y})$).*

In statistics parlance, $x_{[i]}$ denotes the i-th *order statistic* of a sample value. Majorization equivalently says that the order statistic of one sample set dominates another.

In [32], majorization was used in conjunction with another stochastic order to derive a tail bound on the overestimation error of a *icebergs* or heavy-hitters, *i.e.,* objects with a large count/size, on distributed streams of data. We shall discuss this in more detail in Section 5.

We look at one interesting result. Let $X_1, X_2, \cdots, X_n$ be Bernoulli random variables with success probabilities $p_1, p_2, \cdots, p_n$ respectively and $\sum_{i=1}^{n} p_i = \mu$, where $\mu > 0$ is a constant. Gleser [14] proved the following: suppose $\lfloor \mu - 2 \rfloor \leq x \leq \lceil \mu + 2 \rceil$, then the tail probability $\Pr(\sum_{i=1}^{n} X_i \geq x)$ as a function of $p_1, p_2, \cdots, p_n$ is Schur concave. This implies that the bound is maximized when $p_1 = p_2 = \cdots = p_n = \mu/n$, providing a worst case bound.

The result can be extended to various cases. For instance, Merkle and Petrović extended this to the case of independent geometrical and negative binomial random variables [24]. We believe there is room for stronger results, in particular, if the results could be extended to log concave distributions.

4 Negative Association

The concept of negatively associated random variables was first proposed by Joag-Dev *et al.* [19], which is one specific definition of negative dependence between random variables. The definition is as follows [19, Definition 2.1]:

Definition 3. *Random variables $X_1, X_2, \cdots, X_n$ are said to be* negatively associated *if for every pair of disjoint subsets $\mathcal{A}_1, \mathcal{A}_2$ of $\{1, 2, \cdots, n\}$,*

$$Cov\Big(f(X_i, i \in \mathcal{A}_1), g(X_j, j \in \mathcal{A}_2)\Big) \leq 0 \tag{1}$$

whenever f and g are increasing.

A major advantage of negative association over other definitions of negative dependence of random variables is that increasing functions of disjoint sets of negatively associated random variables are also negatively associated, *i.e.,* a closure property is satisfied.

Now, how do negatively associated random variables fit into designing randomized algorithms? The Chernoff bound [25] is a probability bound typically used to bound the error function of a randomized algorithm. However, the chief assumption is that the collection of random variables $X_1, X_2, \cdots, X_n$ must be independently distributed. This is generally not the case in network algorithmics.

Fortunately, the functions of interest on $X_1, X_2, \cdots, X_n$ are often non-decreasing, so the Chernoff bound can be used as an upper bound on function of these random variables, even though $X_1, X_2, \cdots, X_n$ are dependent. One such example is the sum of the random variables, *i.e.,*

$$f(X_1, X_2, \cdots, X_n) = \sum_{i=1}^{n} X_i.$$

Intuition-wise, it's because negative association ensures that the covariance between non-decreasing functions on subsets of the random variables are non-positive. Compare this with independent random variables, where the covariance would be exactly zero. Since the Chernoff bound is a moment bound, we know that non-positive covariance can only decrease variance, so the upper bound must be that of independent random variables. The discussion in the next section and Theorem 9 will make this clearer.

At present, there is no well-developed framework, so proving that a set of random variables are negatively associated is, aside from some special cases, a difficult task.

One useful tool is the Fortuin-Kasteleyn-Ginibre (FKG) inequality [13]. In [9], Dubhashi *et al.* showed how the FKG inequality can be used to prove the negative associativity of some well-known distributions, such as negatively correlated (binary) coins and the permutation distribution. The proofs, however, exploited clever arrangements of the events of a random variable on lattices [9]. It must also be mentioned that this proof technique requires that the set of random variables be discrete random variables.

Unfortunately, at present, it is often difficult to prove negative association despite its wide implications. Though negative association is a very useful property to have, but because proving this property is difficult, one way is to just show a weaker negative dependence, such as, for a particular function f,

$$\mathbb{E}\left[\prod_{i=1}^{m} f(X_i)\right] \leq \prod_{i=1}^{m} \mathbb{E}\left[f(X_i)\right].$$

However, when negative association applies, we can elegantly sidestep more complicated arguments based on martingales and the Azuma-Hoeffding inequality [3, 15].

There has been results proving the negative association of random variables from well-known distributions. Joag-Dev *et al.* [19] list some examples of negatively associated random variables:

(i) *Permutation distribution*: the joint distribution of a random vector $(X_1, X_2, \cdots, X_n)$, $n > 1$, which takes as values all $n!$ permutations of $(x_1, x_2, \cdots, x_n)$, which is a set of n real numbers, with equal probabilities, *i.e.,* $1/n!$. This covers two important cases:

- samples obtained from a finite population via random sampling without replacement, and
- the joint distribution of ranks of a finite random sample from a population.

(ii) *Several canonical multivariate distributions*: examples include the

- multinomial distribution,
- multivariate hypergeometric distribution,

- Dirichlet distribution,
- Dirichlet compound multinomial distribution,
- multinormal distributions having certain covariance matrices, and
- negatively correlated normal random variables.

(iii) *Marginal distributions of the row (column) vectors of a contingency table*: each cell count is an independent random sample taken from subpopulations that were formed by partitioning the population according to the categories of the table.

Aside from these examples, another is the famous *balls-and-bins* model, where n balls are thrown uniformly at random into m bins. If the balls are identical, then the joint distribution of the number of balls in each bin B_i, $i = 1, 2 \cdots, m$, once all the balls were thrown is then equal to the multinomial distribution. Thus, the balls-and-bins model is a generalization of the multinomial distribution, where one can have a probability $p_{i,k}$ to denote the probability ball k lands in bin i. The balls-and-bins model was indirectly listed in Joag-Dev *et al.* [19] as the "convolution of unlike multinomial distributions", but no prove was given.

Dubhashi and Ranjan [10] showed that the random variables B_i, $i = 1, 2 \cdots, m$, are in fact negatively associated. The intuition is clear, though the proof is a little more complicated: since the number of balls n is fixed, if a ball lands in bin i, then there is one less ball that could possibly land in bin j.

The balls-and-bins model arises in various problems in network algorithmics. For instance, the power-of-two-choices heuristic used in load balancing and improving performance of hash tables and Bloom filters is cast as a problem in distributing balls over bins.

4.1 Useful properties

From the above, proving negative association can be challenging. In this section, we list some useful properties of negatively associated random variables. These can be used to simplify proofs of the negative association of a set of random variables.

The following result was presented in [19] without proof. A proof is given here for reference.

Theorem 4. *A pair of continuous random variables X, Y are negatively associated if and only if*

$$\Pr(X \leq x, Y \leq y) \leq \Pr(X \leq x) \Pr(Y \leq y). \tag{2}$$

Proof. If X, Y are negatively associated, then by choosing the indicator functions $\mathbb{I}\{X \leq x\}$ and $\mathbb{I}\{Y \leq y\}$, we have

$$\begin{aligned} \operatorname{Cov}(\mathbb{I}\{X \leq x\}, \mathbb{I}\{Y \leq y\}) &\leq 0 \\ \mathbb{E}[\mathbb{I}\{X \leq x\}\mathbb{I}\{Y \leq y\}] - \mathbb{E}[\mathbb{I}\{X \leq x\}]\mathbb{E}[\mathbb{I}\{Y \leq y\}] &\leq 0 \\ \Pr(X \leq x, Y \leq y) - \Pr(X \leq x) \Pr(Y \leq y) &\leq 0, \end{aligned}$$

and the result follows.

In the other direction, Hoeffding's identity [15] implies

$$\operatorname{Cov}(f(X), g(Y)) = 2 \int_{\mathbb{R}} \int_{\mathbb{R}} \Pr(f(X) \leq u, g(Y) \leq v) - \Pr(f(X) \leq u) \Pr(g(Y) \leq v)\, dx\, dy.$$

However, applying the functions f, g on (2), we get the inequality

$$\Pr(f(X) \leq u, g(Y) \leq v) - \Pr(f(X) \leq u) \Pr(g(Y) \leq v) \leq 0.$$

The result then follows. ■

The result is also equivalent to the following: the pair X, Y is negatively associated if and only if

$$\Pr(Y \leq y \,|\, X \leq x) \leq \Pr(Y \leq y) \text{ or } \Pr(X \leq x \,|\, Y \leq y) \leq \Pr(X \leq x).$$

This equivalence provides much better intuition about negatively associated random variables. Additionally, if the random variables satisfy (2) then they are *negative quadrature dependent* [22].

Note that the theorem does not apply to discrete random variables *i.e.,* . However, in the case of binary random variables, the variables are negatively associated if and only if they are negatively correlated [11].

An important property for use together with Chernoff-type inequalities is the following: let $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_m$ be disjoint subsets of the index set $\{1, 2, \cdots, n\}$ and $f_1, f_2, \cdots, f_m$ be increasing positive functions. Then if the set of random variables $X_1, X_2, \cdots, X_n$ are negatively associated, it implies that

$$\mathbb{E}\left[\prod_{i=1}^{m} f_i(X_j, j \in \mathcal{A}_i)\right] \leq \prod_{i=1}^{m} \mathbb{E}\left[f_i(X_j, j \in \mathcal{A}_i)\right].$$

This means that the function $\prod_{i=1}^{m} f_i(X_j, j \in \mathcal{A}_i)$ is log-concave. Moreover, the it also follows that for $x_i \in \mathbb{R}$, $i = 1, 2, \cdots, n$, and $\mathcal{A}_1, \mathcal{A}_2$ are disjoint subsets of $\{1, 2, \cdots, n\}$,

$$\Pr(X_i \leq x_i, i = 1, 2, \cdots, n) \leq \Pr(X_i \leq x_i, i \in \mathcal{A}_1) \Pr(X_j \leq x_j, j \in \mathcal{A}_2), \tag{3}$$
$$\Pr(X_i > x_i, i = 1, 2, \cdots, n) \leq \Pr(X_i > x_i, i \in \mathcal{A}_1) \Pr(X_j > x_j, j \in \mathcal{A}_2). \tag{4}$$

Several other properties listed in [19] are:

(i) a subset of two or more negatively associated random variables are negatively associated,

(ii) a set of independent random variables are negatively associated,

(iii) increasing functions defined on disjoint subsets of a set of negatively associated random variables are negatively associated,

(iv) the union of independent sets of negatively associated random variables are negatively associated.

5 Convex Ordering and Supermodularity

We begin with the definition of (increasing) convex ordering:

Definition 5. *Let X and Y be random variables with finite means. Then X is less than Y in (increasing) convex order, written $X \leq_{cx} Y$ ($X \leq_{icx} Y$), if $\mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]$ holds for all real (increasing) convex functions f such that the expectations exist.*

We can see how this ordering is useful: the moment generating function of any random variable is an increasing convex function, provided that it exists. Then, an ordering $X \leq_{icx} Y$ implies that the moments of X can be bounded by the moments of Y, which would be useful when using Chernoff-type inequalities (tail bounds).

We will also later on require the *usual stochastic ordering*.

Definition 6. *A random variable (or random vector) X is stochastically less than or equal to a random variable (or random vector) Y, denoted as $X \leq_{st} Y$, if and only if $\mathbb{E}[\phi(X)] \leq \mathbb{E}[\phi(Y)]$ for all increasing functions ϕ such that the expectations exist.*

Supermodular functions (also know as L-superadditive functions [4]) are defined as follows:

Definition 7. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is called supermodular if*

$$f(\mathbf{x} \vee \mathbf{y}) + f(\mathbf{x} \wedge \mathbf{y}) \geq f(\mathbf{x}) + f(\mathbf{y})$$

where

$$\mathbf{x} \vee \mathbf{y} := [\max(x_1, y_1), \max(x_1, y_1), \cdots, \max(x_n, y_n)],$$
$$\mathbf{x} \wedge \mathbf{y} := [\min(x_1, y_1), \min(x_1, y_1), \cdots, \min(x_n, y_n)].$$

A function f is submodular if and only if $-f$ is supermodular.

Supermodular functions are often defined on lattices. Naturally, a partial order can be defined on supermodular functions:

Definition 8. *A random vector $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ is said to be smaller than a random vector $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_n)$ in the supermodular order, denoted by $\boldsymbol{X} \leq_{sm} \boldsymbol{Y}$ if $\mathbb{E}f(\boldsymbol{X}) \leq \mathbb{E}f(\boldsymbol{Y})$ for all supermodular functions f for which expectations exist.*

A general result, by Christofides and Vaggelatou [7, Theorem 1(b)], proved an ordering for all supermodular functions with negatively associated random variables. Supermodular ordering first appeared in [30].

Theorem 9 (Supermodular ordering and negative association)**.** *Let $X_1, X_2, \cdots, X_n$ be a collection of negatively associated random variables and $Y_1, Y_2, \cdots, Y_n$ be independent random variables where each Y_i possesses the same marginal distribution as X_i, $\forall i$. Then,*

$$(X_1, X_2, \cdots, X_n) \leq_{sm} (Y_1, Y_2, \cdots, Y_n).$$

The fact that a composition between an increasing and convex function $f : \mathbb{R}^n \to \mathbb{R}$ and monotone and supermodular function $\phi : \mathbb{R}^n \to \mathbb{R}$ results in $f \circ \phi$ being supermodular results in the following corollary [7, Corollary 1(b)]:

Corollary 10. *Let $X_1, X_2, \cdots, X_n$ be a collection of negatively associated random variables and $Y_1, Y_2, \cdots, Y_n$ be independent random variables where each Y_i possesses the same marginal distribution as X_i, $\forall i$. Then,*

$$\phi(X_1, X_2, \cdots, X_n) \leq_{icx} \phi(Y_1, Y_2, \cdots, Y_n),$$

for every ϕ monotone and supermodular.

Let us consider a simple example. The linear function

$$f(x) = \sum_i a_i x_i, \tag{5}$$

for all $a_i > 0$ is both increasing and supermodular (note that if there exists for some i, $a_i < 0$, then the function is no longer increasing and supermodular). This function was encountered in the design of an SRAM-DRAM hybrid statistic counter architecture for counting traffic flow sizes [33]. We thus have a simpler alternative proof of [33, Theorem 3]:

Theorem 11. *Let $a_1, a_2, \cdots, a_n$ be constants, with $a_i > 0$, $\forall i$. Let $X_1, X_2, \cdots, X_n$ be real-valued negatively associated random variables. Let $Y_1, Y_2, \cdots, Y_n$ be independent random variables where each Y_i possesses the same marginal distribution as X_i, $\forall i$. Then,*

$$\sum_{i=1}^{n} a_i X_i \leq_{cx} \sum_{i=1}^{n} a_i Y_i.$$

Proof. Note that $\mathbb{E}[\sum_i a_i X_i] = \mathbb{E}[\sum_i a_i Y_i]$. Then, convex ordering is obtained instead of an increasing convex ordering. The result is just a consequence of the monotone increasing nature and supermodularity of the function (5). ■

Remark 12. *Note that it is not enough for X_is to be negatively correlated. Theorem 11 asserts that X_is are negatively correlated for all convex functions f. This can be seen in the proof of [7, Theorem 1(b)], where we need $Cov(f'_+(X_1 + t), \mathbb{I}\{X_2 > t\}) \leq 0$ to hold for all t.*

From this result, we immediately get the following as a corollary:

Corollary 13. *Assume the same notation as above. Let $X_1, X_2, \cdots, X_n$ be a sample without replacement and $Y_1, Y_2, \cdots, Y_n$ be a sample with replacement from a population $S = \{c_1, c_2, \cdots, c_N\}$, where $N > n$. Then,*

$$\sum_{i=1}^{n} a_i X_i \leq_{cx} \sum_{i=1}^{n} a_i Y_i.$$

Proof. It is well-known that since $X_1, X_2, \cdots, X_n$ are a set of samples without replacement, they are negatively associated [19]. The result follows from a direct application of Theorem 11. ■

Moreover, Theorem 11 can be extended to show

$$\max_{1\leq k\leq n} \sum_{i=1}^{k} a_i X_i \leq_{\text{icx}} \max_{1\leq k\leq n} \sum_{i=1}^{k} a_i Y_i.$$

Such a result can be used for the worse case tail analysis. Note that the ordering becomes a convex ordering *i.e.,*

$$\max_{1\leq k\leq n} \sum_{i=1}^{k} a_i X_i \leq_{\text{cx}} \max_{1\leq k\leq n} \sum_{i=1}^{k} a_i Y_i,$$

if $\mathbb{E}[X_i] = \mathbb{E}[Y_i]$ for all i. Other examples of useful functions are the order statistic functions, for instance, the first and last order statistics, which are supermodular and submodular respectively.

Now, a set of random variables $X_1, X_2, \cdots, X_n$ are *exchangeable* if their joint distribution is invariant under permutation. Suppose X_i and Y_i are exchangeable for $i = 1, 2, \cdots, n$. Then the following holds

Theorem 14. *Let $a_1, a_2, \cdots, a_n$ be constants, with $a_i > 0$, $\forall i$. Let $X_1, X_2, \cdots, X_n$ be real-valued exchangeable and negatively associated random variables. Let $Y_1, Y_2, \cdots, Y_n$ be real-valued exchangeable independent random variables where each Y_i possesses the same marginal distribution as X_i, $\forall i$. Then, for every $\mathcal{S} \subseteq \{1, 2, \cdots, n\}$,*

$$\sum_{i\in\mathcal{S}} a_i X_i \leq_{icx} \sum_{i\in\mathcal{S}} a_i Y_i$$

and

$$\max_{1\leq|\mathcal{S}|\leq n} \sum_{i\in\mathcal{S}} a_i X_i \leq_{icx} \max_{1\leq|\mathcal{S}|\leq n} \sum_{i\in\mathcal{S}} a_i Y_i.$$

Proof. We can construct any subset $\mathcal{S}$ as follows. Suppose $|\mathcal{S}| = k$. Consider a permutation of X_i and Y_i, $X_{\sigma(i)}$ and $Y_\sigma(i)$ respectively for $i = 1, 2, \cdots, n$. Then, indices belonging to $\mathcal{S}$ can be chosen simply by choosing the appropriate permutation map σ and then sampling the first k indices. Since the random variables are exchangeable, the joint distribution of these random variables remain unchanged, so they remain negatively associated. The theorem then follows by applying Theorem 11. ■

Ideas from majorization can then be combined with supermodularity as well. Here, we present a result found in [32] on the design of a distributed algorithm for detecting global icebergs in networks.

Theorem 15 ([32])**.** *Let f be a convex function and $X_1, X_2, \cdots, X_n$ be non-negative valued exchangeable random variables. Then, for two real vectors $\mathbf{a}$ and $\mathbf{b}$ with the relation $\mathbf{a} \leq_M \mathbf{b}$ implies*

$$\sum_{i=1}^{n} f(a_i) X_i \leq_{icx} \sum_{i=1}^{n} f(b_i) X_i.$$

5.1 Proof recipe

Suppose we know that a set of random variables $X_1, X_2, \cdots, X_n$ are negatively associated. The general result of Christofides and Vaggelatou suggests a recipe on proving tail bounds on a function of disjoint subsets of X_is: simply focus on proving the supermodularity (submodularity) of the function f.

To do so, we require a closely related concept called *(strictly) increasing differences*. For any function $f : \mathbb{R}^n \to \mathbb{R}$, let any pair of indices $i, j \in \{1, 2, \cdots, n\}$, and any vector

$$\hat{x}_{i,j} = (x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_{j-1}, x_{j+1}, \cdots, x_n) \in \mathbb{R}^{n-2}$$

that is an exclusion of entries i, j, and

$$f_{\hat{x}_{i,j}}(x'_i, x'_j) := f(x_1, \cdots, x_{i-1}, x'_i, x_{i+1}, \cdots, x_{j-1}, x'_j, x_{j+1}, \cdots, x_n).$$

A function f then has (strictly) increasing differences if for any pair of distinct indices i, j and any vector $\hat{x}_{i,j}$, with $x_i \leq x'_i (x_i < x'_i)$ and $x_j \leq x'_j (x_j < x'_j)$,

$$f_{\hat{x}_{i,j}}(x_i, x_j) - f_{\hat{x}_{i,j}}(x_i, x'_j)(<) \leq f_{\hat{x}_{i,j}}(x'_i, x_j) - f_{\hat{x}_{i,j}}(x'_i, x'_j).$$

The definition can be extended to the integer domain as well. For instance, we can see that the maximum bin load in the balls and bins model is an example of an increasing differences function.

Then, following result would be useful in proving supermodularity of a function:

Theorem 16. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is (strictly) supermodular if and only if f has (strictly) increasing differences.*

Moreover, if the supermodular function is twice differentiable, then to prove supermodularity one needs to show

$$\frac{\partial^2 \phi(\mathbf{x})}{\partial x_i \partial x_j} \geq 0,$$

for $\mathbf{x} \in \mathbb{R}^n$ and all $i \neq j$.

Example supermodular and increasing functions:

- p-norms: $\|\mathbf{x}\|_p = \left(\sum_i x_i^p\right)^{1/p}$,
- first order statistic [4]: $f(\mathbf{x}) = \max_{1 \leq i \leq n} x_i$,
- product function over $\mathbb{R}^n \times \mathbb{R}^n$: $f(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i y_i$,
- Cobb Douglas function: $f(\mathbf{x}) = \prod_{i=1}^n x_i^{\alpha_i}$ on the set $\{\mathbf{x} \,|\, \mathbf{x} \succeq 0\}$ ($\succeq$ denotes element-wise non-negativity), for $\alpha_i \geq 0$,
- minimization: if $f_i(z)$ is increasing on $\mathbb{R}$ for $i = 1, 2, \cdots, n$ then

  $$f(\mathbf{x}) = \min_{i \in \{1,2,\cdots,n\}} f_i(x_i)$$

  is supermodular on $\mathbb{R}^n$.

6 Log-Concave Distributions

Here, we briefly mention about log-concave distributions. A more complete survey of the topic is found in [2] and [28].

A continuous distribution is *log-concave* if its domain is a convex set and its probability density function $f : \mathbb{R}^n \to [0, 1]$ satisfies

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \geq f(\mathbf{x})^\lambda f(\mathbf{y})^{1-\lambda} \tag{6}$$

for all $\mathbf{x}, \mathbf{y} \in \operatorname{dom} f$ and $0 < \lambda < 1$. We can see this is equivalent to

$$\log f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \geq \lambda \log f(\mathbf{x}) + (1 - \lambda) \log f(\mathbf{y})$$

if f is strictly positive. Examples include well-known distributions such as the normal distribution, Wishart distribution, Dirichlet distribution and Beta distribution.

In network algorithmics, for counting applications, we frequently deal with integer-valued (discrete) random variables. The definition can be extended for integer-valued random variables with modifications. In the discrete case, log-concave distributions are defined as probability mass functions satisfying

$$p_x^2 \geq p_{x-1} p_{x+1}$$

where $p_x := \Pr(X = x)$ for all $x \in \mathbb{Z}$ [20]. In the case of [16, 17, Theorem 1], it was shown that the independent samples from the Binomial distribution is log-concave. Other examples include [2, 20]:

(i) Bernoulli trial,

(ii) the binomial distribution,

(iii) the Poisson distribution,

(iv) the geometric distribution, and

(v) the negative binomial distribution.

These canonical distributions are often encountered in counting sketch applications.

Let Ω denote the support of the probability density function of a continuous random variable X.

Definition 17. *A probability density $f(x)$ is unimodal if there exists a mode $m \in \Omega$ such that $f(x) \leq f(y)$ for all $x \leq y \leq m$ or for all $m \leq y \leq x$. $f(x)$ is strongly unimodal if the convolution of f with any unimodal g is unimodal.*

A surprising result by Ibragimov [18] is the following:

Theorem 18. *A random variable X is distributed according to a log-concave distribution if and only if its density function $f(x)$ is strongly unimodal.*

This is useful, because one can check if a distribution is strongly unimodal more readily than using the direct definition of a log-concave distribution.

Though Ibragimov's result applies to continuous random variables, by applying the definition of log-concavity for discrete random variables, the results can be extended to log-concave distributions in the discrete case [20, Theorem 3].

Log-concave distributions have several desirable properties for applications to randomized algorithms:

(i) the tail of a log-concave distribution is *subexponential*, *i.e.,* it decays faster than the tail of an exponential distribution,

(ii) the convolution of log-concave distributions is log-concave *i.e.,* the family is closed under convolution.

The first property is clearly useful in bounding the worst-case probability of "bad" events.

The second is useful in proving more complex distributions are log-concave. For instance, the sum of unlike Bernoulli trials, *i.e.,* the distribution of Bernoulli trials, each with parameter p_i, $i = 1, 2, \cdots, n$, has a distribution that is the convolution of the distribution of the trials. Since a Bernoulli trial has a log-concave distribution, then the sum of unlike Bernoulli trials also has a log-concave distribution. The sum of random variables is often encountered in many problems in network algorithmics.

6.1 Efron's monotonicity theorem

The importance of log-concave distributions is chiefly due to *Efron's monotonicity theorem* [12]. The theorem states the following:

Theorem 19. *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ where f is coordinate-wise non-decreasing and let*

$$g(z) := \mathbb{E}\left[f(X_1, X_2, \cdots, X_n) \,\middle|\, \sum_{i=1}^{n} X_i = z \right], \tag{7}$$

where $X_1, X_2, \cdots, X_n$ are independent and log-concave. Then g is non-decreasing.

Note that this result also holds for integer-valued (discrete) random variables with a log-concave distribution.

An interesting consequence of Theorem 19 is that the joint conditional distribution of $X_1, X_2, \cdots, X_n$ given $\sum_{i=1}^{n} X_i$ is *negatively associated* (almost surely) [19, Theorem 2.8]. For instance, this means samples with replacement from a finite population, conditioned on a total sampling budget are negatively associated. From the previous section, negative association implies some useful properties that can be used to bound worse case events, for example, as in Theorem 14.

Moreover, by the definition of the usual stochastic order, Efron's theorem implies

$$\left(X_1, X_2, \cdots, X_n \,\middle|\, \sum_{i=1}^{n} X_i = s\right) \leq_{\text{st}} \left(X_1, X_2, \cdots, X_n \,\middle|\, \sum_{i=1}^{n} X_i = t\right) \tag{8}$$

for $t > s$ if $X_1, X_2, \cdots, X_n$ are independent samples from log-concave distributions.

One regularly encountered bound is a bound we term the *times-2 bound*: for any coordinate-wise non-decreasing function f, and two sets of random variables $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_n$,

$$\mathbb{E}[f(X_1, X_2, \cdots, X_n)] \leq 2\mathbb{E}[f(Y_1, Y_2, \cdots, Y_n)], \tag{9}$$

where Y_i has the same marginal distribution as X_i for all i, yet Y_is are independent.

For the balls-into-bins model with m identical balls and n bins, this was proven rigorously by Mitzenmacher and Upfal [25]: the X_is denote the number of balls in a bin after all the balls have been thrown (which are clearly dependent), while the Y_is denote the independent Poisson random variables with rate m/n.

We present a generalized version of the bound that applies to log-concave distributions.

Theorem 20 (Times-2 Bound)**.** *Let $Y_1, Y_2, \cdots, Y_n$ be a set of independent log-concavely distributed random variables, and let the support set of the distribution of the random variable $\sum_{i=1}^{n} Y_i$ be Ω. Suppose*

(i) there exists a proper subset $\Omega' \subset \Omega$, which is a restricted version of Ω, starting from τ to the end of the maximum support (can be ∞), and

(ii) τ is less than the median of $\sum_{i=1}^{n} Y_i$.

Let $X_1, X_2, \cdots, X_n$ be a set of dependent random variables such that

$$\mu(X_1, X_2, \cdots, X_n) = \mu\left(Y_1, Y_2, \cdots, Y_n \,\middle|\, \sum_{i=1}^{n} Y_i = \tau\right),$$

where $\mu(Z)$ is the distribution of a random variable or vector Z. Then, for any coordinate-wise non-decreasing function $f(x_1, x_2, \cdots, x_n)$,

$$\mathbb{E}[f(X_1, X_2, \cdots, X_n)] \leq 2\mathbb{E}[f(Y_1, Y_2, \cdots, Y_n)]. \tag{10}$$

Proof. Here we derive the result for the case when both X_i and Y_i are discrete for all i. The proof applies similarly when both X_i and Y_i are continuous with some modifications. The proof follows an outline from the proof of [16, Theorem 1] and [25, p. 121]. For any coordinate-wise non-decreasing function $f(x_1, x_2, \cdots, x_n)$,

$$\begin{aligned}
\mathbb{E}[f(Y_1, Y_2, \cdots, Y_n)] &= \sum_{\ell \in \Omega} \mathbb{E}\left[f(Y_1, Y_2, \cdots, Y_n) \,\middle|\, \sum_{i=1}^{n} Y_i = \ell\right] \Pr\left(\sum_{i=1}^{n} Y_i = \ell\right) \\
&\geq \sum_{\ell \in \Omega'} \mathbb{E}\left[f(Y_1, Y_2, \cdots, Y_n) \,\middle|\, \sum_{i=1}^{n} Y_i = \ell\right] \Pr\left(\sum_{i=1}^{n} Y_i = \ell\right) \\
&\geq \sum_{\ell \in \Omega'} \mathbb{E}\left[f(Y_1, Y_2, \cdots, Y_n) \,\middle|\, \sum_{i=1}^{n} Y_i = \tau\right] \Pr\left(\sum_{i=1}^{n} Y_i = \ell\right) \qquad (11) \\
&= \mathbb{E}\left[f(Y_1, Y_2, \cdots, Y_n) \,\middle|\, \sum_{i=1}^{n} Y_i = \tau\right] \sum_{\ell \in \Omega'} \Pr\left(\sum_{i=1}^{n} Y_i = \ell\right) \\
&= \mathbb{E}\left[f(X_1, X_2, \cdots, X_n)\right] \Pr\left(\sum_{i=1}^{n} Y_i \geq \tau\right) \\
&\geq \frac{1}{2}\mathbb{E}\left[f(X_1, X_2, \cdots, X_n)\right]. \qquad (12)
\end{aligned}$$

Inequality (11) follows from the implication of Efron's theorem, that is (8) and Condition (i). Inequality (12) follows since Condition (ii) implies

$$\Pr\left(\sum_{i=1}^{n} Y_i \geq \tau\right) \geq \frac{1}{2}.$$

■

Remark 21. *Now if $X_1, X_2, \cdots, X_n$ has a joint distribution such that*

$$\mu(X_1, X_2, \cdots, X_n) = \mu\left(Y_1, Y_2, \cdots, Y_n \,\middle|\, \sum_{i=1}^{n} Y_i = \tau\right)$$

and since $Y_1, Y_2, \cdots, Y_n$ given $\sum_{i=1}^{n} Y_i = \tau$ are negatively associated (almost surely), by implication, $X_1, X_2, \cdots, X_n$ are negatively associated (almost surely). The proof, which we omit here, is via contradiction.

We can say a little more about the result. In Condition (ii), it is stated that τ is less than the median of $\sum_{i=1}^{n} Y_i$. Often, what one would like to do is to set τ to be the mean of $\sum_{i=1}^{n} Y_i$. We know that log-concave distributions are strongly unimodal, so the distribution of $\sum_{i=1}^{n} Y_i$ is strongly unimodal. For the condition to be satisfied, we require that the mean must be less than or equal to the median.

The result has interesting implications: we can use a log-concavely distributed random variables to bound a set of (negatively) dependent random variables. In the balls-and-bins model, the chosen distribution for the Y_is are the Poisson [25] and Binomial [16] random variables. Our result shows that a wider class of distributions can be applied, so as long as the conditions of Theorem 20 are satisfied.

With this result, we can now derive simple bounds. One trick is to observe that the indicator function is a coordinate-wise non-decreasing function, so we can set

$$f(x_1, x_2, \cdots, x_n) = \mathbb{I}\Big\{g(x_1, x_2, \cdots, x_n) > c\Big\},$$

for some constant c and some function g. Then, assuming the conditions of Theorem 20 are satisfied, this gives us

$$\Pr\Big(g(X_1, X_2, \cdots, X_n) > c\Big) \leq 2\Pr\Big(g(Y_1, Y_2, \cdots, Y_n) > c\Big).$$

Since a tail bound is more readily derived for the Y_is since they are independent, we know that the bound on $g(X_1, X_2, \cdots, X_n)$ is at most twice the tail bound probability on the Y_is.

As an example, in the balls-and-bins model, we can set

$$g(x_1, x_2, \cdots, x_n) = \max_{1 \leq i \leq n} x_i,$$

that is, the maximum load over all bins. We then choose Y_is from the Poisson distribution with rate $\alpha = m/n$. Work by Kimber [21] showed that

$$\Pr\left(\max_{1 \leq i \leq n} Y_i > \frac{\log n}{\log \log n} + 1\right) \to 0, \tag{13}$$

for fixed α as $n \to \infty$, so with high probability, the maximum load of a bin is $O(\log n / \log \log n)$, confirming the results in literature.

As an aside, there is no known result on the necessary condition for (7) to be coordinate-wise non-decreasing with respect to z. However, by following the proof of Efron's theorem, it is clear that the properties of the distribution must be preserved under convolution, since adding X_{n+1} to a sequence $X_1, X_2, \cdots, X_n$ should preserve the joint distribution in order for the condition to hold. For instance, log-concave distributions satisfy this condition because they are closed under convolution.

7 Conclusion

Stochastic orders are excellent tools to aid in the design of randomized algorithms. We see this through their application in simplifying proofs for bounding the errors of these algorithms. Though our focus is on network algorithmics, stochastic orders can be applied in a broader context. We hope that through our paper, researchers are made aware of the benefits of exploiting these tools.

Acknowledgment

This work is supported in part by US NSF through grant CNS 1302197 and the Australian Research Council (ARC) grant DP110103505..

References

[1] N. Alon and J. H. Spencer. *The Probabilistic Method*. John Wiley and Sons, 2nd edition, 2004.

[2] M. Y. An. Log-concave probability distributions: Theory and statistical testing. Technical report, Duke University, November 1995.

[3] K. Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 19(3):357–367, 1967.

[4] H. W. Block, W. S. Griffith, and T. H. Savits. L-superadditive structure functions. *Adv. Appl. Prob.*, 21:919–929, 1989.

[5] C.-S. Chang, D.-S. Lee, and Y.-S. Jou. Load balanced Birkhoff-von Neumann switches, part I: One-stage buffering. *Comput. Commun.*, 25(6):611–622, April 2002.

[6] C.-S. Chang, D.-S. Lee, and Y.-S. Jou. Load balanced Birkhoff-von Neumann switches, part II: Multi-stage buffering. *Comput. Commun.*, 25(6):623–634, April 2002.

[7] T. Christofides and E. Vaggelatou. A connection between supermodular ordering and positive/negative association. *Journal of Multivariate Analysis*, 88(1):138–151, January 2004.

[8] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*. Applications of mathematics. Springer, New York, Berlin, Heidelberg, 1998.

[9] D. Dubhashi, V. Priebe, and D. Ranjan. Negative dependence through the FKG inequality. Technical report, University of Aarhus, July 1996.

[10] D. Dubhashi and D. Ranjan. Balls and bins: A study in negative dependence. Technical report, University of Aarhus, August 1996.

[11] D. Dubhashi and D. Ranjan. Balls and bins: A study in negative dependence. *Random Structures and Algorithms*, 13:99–124, 1996.

[12] B. Efron. Increasing properties of Polya frequency function. *Ann. Math. Statist.*, 36(1):272–279, 1965.

[13] C. M. Fortuin, J. Ginibre, and P. W. Kasteleyn. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22(2):89–103, 1971.

[14] L. J. Gleser. On the distribution of the number of successes in independent trials. *Ann. Probab.*, 3(1):182–188, 1975.

[15] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Am. Statist. Assoc.*, 58(301):13–30, 1963.

[16] N. Hua, B. Lin, J. Xu, and H. Zhao. BRICK: A novel exact active statistics counter architecture. In *ANCS '08*, November 2008.

[17] N. Hua, B. Lin, J. Xu, and H. Zhao. BRICK: A novel exact active statistics counter architecture. *IEEE/ACM Trans. Networking*, 19(3):670–682, June 2011.

[18] I. A. Ibragimov. On the composition of unimodal distributions. *Theory Probab. Appl.*, 1(2):255–260, 1965.

[19] K. Joag-Dev and F. Proschan. Negative association of random variables with applications. *Ann. Statist.*, 11(1):286–295, 1983.

[20] J. Keilson and H. Gerber. Some results for discrete unimodality. *J. Amer. Statist. Assoc.*, 66(334), June 1971.

[21] A. C. Kimber. A note on Poisson maxima. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 63:551–552, 1983.

[22] E. L. Lehmann. Some concepts of dependence. *Ann. Math. Statist.*, 43(5):1137–1153, 1966.

[23] A. W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and its Applications*. Academic, 1979.

[24] M. Merkle and L. Petrović. Inequalities for sums of independent geometrical random variables. *Aequ. Math.*, 54:173–180, 1997.

[25] M. Mitzenmacher and E. Upfal. *Probability and Computing*. Cambridge University Press, 2005.

[26] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

[27] A. Müller and D. Stoyan. *Comparison methods for stochastic models and risks*. Wiley series in probability and statistics. John Wiley, 2002.

[28] A. Saumard and J. A. Wellner. Log-concavity and strong log-concavity: A review. *Statist. Surv.*, 8:45–114, 2014.

[29] M. Shaked and G. J. Shanthikumar. *Stochastic orders and their applications*. Probability and mathematical statistics. Academic Press, 2nd edition, 2007.

[30] R. Szekli, R. L. Disney, and S. Hur. $MR/GI/1$ queues with positively correlated arrival stream. *J. Appl. Probab.*, 31(2):397–408, June 1994.

[31] G. Varghese. *Network Algorithmics: An Interdisciplinary Approach to Designing Fast Networked Devices*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.

[32] H. Zhao, A. Lall, M. Ogihara, and J. Xu. Global iceberg detection over distributed data streams. In *IEEE International Conference on Data Engineering (ICDE)*, March 2010.

[33] H. C. Zhao, H. Wang, B. Lin, and J. J. Xu. Design and performance analysis of a dram-based statistics counter array architecture. In *Proc. of ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS)*, October 2009.