A Calculus for Information-Driven Networks

Kui Wu
Computer Science Department
University of Victoria
BC, Canada V8W 3P6

Yuming Jiang, Guoqiang Hu
Q2S Center of Excellence
Norwegian University of Science and Technology
Trondheim, Norway

***Abstract*—Information-driven networks include a large category of networking systems, where network nodes are aware of information delivered and thus can not only forward data packets but may also perform information processing. In many situations, the quality of service (QoS) in information-driven networks is provisioned with the redundancy in information. Traditional performance models generally adopt evaluation measures suitable for packet-oriented service guarantee, such as packet delay, throughput, and packet loss rate. These performance measures, however, do not align well with the actual need of information-driven networks. New performance measures and models for information-driven networks, despite their importance, have been mainly blank, largely because information processing is clearly application dependent and cannot be easily captured within a generic framework. To fill the vacancy, we develop a new performance evaluation framework particularly tailored for information-driven networks, based on the recent development of stochastic network calculus. Particularly, our model captures the information processing and the QoS guarantee with respect to stochastic information delivery rates, which have never been formally modeled before. This analytical model is very useful in deriving theoretical performance bounds for a large body of systems where QoS is stochastically guaranteed with a certain level of information delivery.**

***Index Terms*—Network Calculus, Information-Driven Networks, Performance Modeling**

I. Introduction

Although computer networks in general are purposed for information delivery, most existing network architectures like the Internet are actually not information driven in the sense that network nodes (e.g., routers and switchers) only care about packets instead of the information inside. As a common principle, network nodes as well as the whole network system are designed to support quality of service (QoS) with respect to packet-oriented service measures such as bounded packet delay and promised data throughput. To achieve this, QoS provisioning mechanisms [3] have been proposed and used in the Internet. It has been observed that on the one hand QoS provisioning mechanisms provide certain service guarantee to privileged data traffic; on the other hand they largely increase the system complexity and incur a heavy burden on network nodes. Many emerging network systems, for example, wireless sensor networks, consist of nodes with very limited computational capability and thus do not have the luxury to accommodate complex QoS mechanisms. Nevertheless, QoS is important in any means. For instance, in a patient-monitoring system or a fire alarm system with wireless sensor networks, we certainly require important information like abnormal heart beats or high temperature readings to be delivered correctly to a monitoring center. The dilemma we face is to guarantee QoS maybe without any underlying promise from network nodes on timely per packet delivery.

The traditional meaning of QoS, e.g., for guaranteed per packet delivery and end-to-end delay, is actually an overkill, since all we care is information. The traditional solutions focusing on packets instead of the information inside have the historical reason: the network protocol stack is layered and network protocols should not mix up with application-layer information. With the emergence of new technologies such as wireless sensor networks, however, the layering principle is not necessarily a rule of thumb, and the redundancy in the information sources should be utilized in network protocol design. A network node may not be purely a data forwarding device. Instead, it may become aware of information forwarded and is able to perform information processing whenever necessary. The ultimate goal of the whole network system is no longer to guarantee service for individual packets, but to guarantee a certain amount of information to be successfully transported. We call this type of networking systems *information-driven networks*. Typical examples include wireless sensor networks with directed data diffusion [9], distributed content sharing over peer-to-peer networks [2], [12], and networks using network coding [1], [22].

Making the network to be information driven opens special opportunities for QoS provisioning, e.g., in environment where the network is subject to high packet losses or network nodes are stringently constrained by computational power and limited bandwidth. In applications where data exhibit spatial and/or temporal correlation, it is unnecessary to provide reliable transmission for each individual packet. Instead, QoS is guaranteed as long as required information can be obtained as sure (i.e., with a very high probability).

Example 1: Assume that a wireless sensor network includes six sensor nodes and one processing center, also called the sink node, as shown in Figure 1. Four sensors at the bottom of the figure monitor the environment and periodically send out measurement data like temperature and humidity readings. Two sensors in the middle of the figure are used as data relay to the sink. Wireless links are generally subject to a high loss rate in wireless sensor networks, so we assume that the average packet loss rate is 25% for each wireless link. Without considering information, we treat the network as purely a data delivery system like the Internet. In this case, we

978-1-4244-3876-1/09/$25.00 ©2009 IEEE

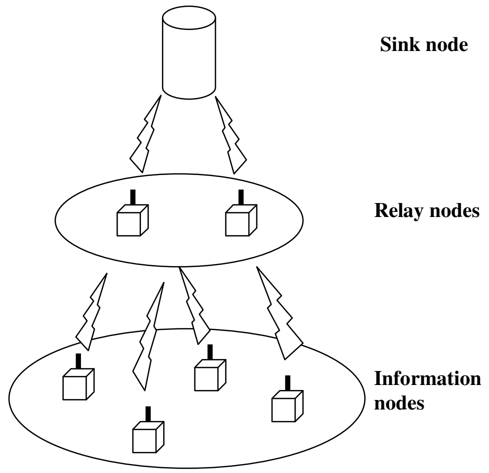

Fig. 1. A simple example of wireless sensor networks

need to make sure that each data packet is correctly delivered from the source to the sink with a high probability. If we set this probability to be no smaller than 96%, we need about 24 transmissions in total (calculated with two retransmissions each link to guarantee the high probability of correct end-to-end packet delivery). In contrast, if relay nodes know that the information from the four source sensors is highly correlated and if the information is considered to be delivered as long as at least one packet from the sources is received by the sink, eight transmissions (e.g., without any retransmissions) can guarantee that the information is delivered with a probability no smaller than 96%.

The above example clearly illustrates the necessity of taking information into consideration. Yet, several difficulties need to address even in the very simple example. First, how can we capture and model the correlation at the information sources? In addition, the correlation may change over time. How can we capture the dynamic changes in a timely fashion? Second, the above example only considers the correlation at the information sources, how can we perform information processing at intermediate relay nodes for better QoS provisioning and resource saving? Third, the use of application-layer information in network protocols has changed the fundamental design principle of current Internet architecture, where the network is considered as a packet transportation tool and the service guarantee is promised for individual packets. This fundamental change renders traditional performance modeling and evaluation approaches invalid for information-driven networks. For instance, network throughput in terms of number of bits per time unit and end-to-end packet delay are no longer good measures and new metrics should be used to align with the need of information-driven networks. What should be a good model for performance evaluation and resource scheduling for information-driven networks?

During the last several years, there are substantial research efforts devoted to tackling the first two difficulties. Particularly, the spatial and temporal correlations of information have been studied and utilized in network scheduling and resource saving in wireless sensor networks [16], [18], [20]; information redundancy has been exploited to help load balancing and improve fault tolerance in peer-to-peer content sharing systems [2], [12]. Regarding the third challenge, to the best of our knowledge, the only attempts to accommodate information processing in performance modeling are the work in [8], [19]. Nevertheless, information processing is simply modeled with a scaling function in [8], [19] and the information embedded in data packets has not been modeled, let alone utilized. In this sense, the performance models in [8], [19] are not really information-driven models. A systematic performance modeling framework suitable for information-driven networks still remains largely open.

In this paper, we develop the first-of-the-kind analytical model suitable for performance study of information-driven networks, based on the recent development of stochastic network calculus. Particularly, our model captures the information processing and the QoS guarantee with respect to stochastic information delivery, which have never been formally modeled before. This analytical model is very useful in deriving theoretical performance bounds for information-driven networks.

II. Background of Stochastic Network Calculus

A. Notation

We first introduce the notation and key concepts of stochastic network calculus [10], [11], [15]. Throughout this paper, we assume that all arrival curves and service curves are non-negative and wide-sense increasing functions. Conventionally, $A(t)$ and $A^*(t)$ are used to denote the *cumulative* traffic that arrives and departures in time interval $(0,t]$, respectively, and $S(t)$ is used to denote the cumulative amount of service provided by the system in time interval $(0,t]$. For any $0 \leq s \leq t$, $A(s,t) \equiv A(t) - A(s)$, $A^*(s,t) \equiv A^*(t) - A^*(s)$, and $S(s,t) \equiv S(t) - S(s)$. By default, $A(0) = A^*(0) = S(0) = 0$.

We denote by $\mathcal{F}$ the set of non-negative wide-sense increasing functions, i.e.,

$$\mathcal{F} = \{f(\cdot) : \forall 0 \leq x \leq y, 0 \leq f(x) \leq f(y)\},$$

and by $\bar{\mathcal{F}}$ the set of non-negative wide-sense decreasing functions, i.e.,

$$\bar{\mathcal{F}} = \{f(\cdot) : \forall 0 \leq x \leq y, 0 \leq f(y) \leq f(x)\}.$$

For any random variable X, its distribution function, denoted by $F_X \equiv Prob\{X \leq x\}$, belongs to $\mathcal{F}$, and its complementary distribution function, denoted by $\bar{F}_X \equiv Prob\{X > x\}$, belongs to $\bar{\mathcal{F}}$.

B. Operators

The following operations defined under the $(\min, +)$ algebra [4], [7], [14] will be used in this paper:

- The $(\min, +)$ *convolution* of functions f and g is

$$(f \otimes g)(t) = \inf_{0 \leq s \leq t} \{f(s) + g(t-s)\}. \qquad (1)$$

- The $(\min, +)$ *deconvolution* of functions f and g is

$$(f \oslash g)(t) = \sup_{s \geq 0} \{f(t+s) - g(s)\}. \qquad (2)$$

- The $(\min, +)$ *inf-sum* of functions f and g is

$$(f \odot g)(t) = \inf_{s \geq 0}\{f(t+s) + g(s)\}. \quad (3)$$

C. Performance Measures, Traffic and Server Models

The following measures are of interest in service guarantee analysis under network calculus:

- The backlog $\mathcal{B}(t)$ in the system at time t is defined as:

$$\mathcal{B}(t) = A(t) - A^*(t). \quad (4)$$

- The delay $\mathcal{D}(t)$ at time t is defined as:

$$\mathcal{D}(t) = \inf\{\tau \geq 0 : A(t) \leq A^*(t+\tau)\}. \quad (5)$$

Stochastic arrival curve and stochastic service curve are core concepts in stochastic network calculus with the former for traffic modeling and the latter for server modeling. It is worth noting that the deterministic traffic arrival curve and the deterministic service curve under the (deterministic) network calculus are a special case of their corresponding stochastic definition. In the literature, there are different definitions of stochastic arrival curve and stochastic service curve [10], [11] such as:

Definition 1: A flow $A(t)$ is said to have a *maximum-virtual-backlog-centric stochastic arrival curve* $\alpha \in \mathcal{F}$ with bounding function $f \in \bar{\mathcal{F}}$, denoted by $A \sim_{m.b.c.} < f, \alpha >$, iff for all $t \geq 0$ and all $x \geq 0$, there holds [10], [11]

$$Prob\{\sup_{0\leq s\leq t}\sup_{0\leq u\leq s}[A(u,s) - \alpha(s-u)] > x\} \leq f(x). \quad (6)$$

Definition 2: A server is said to provide a flow $A(t)$ with a *stochastic service curve* $\beta \in \mathcal{F}$ with bounding function $g \in \bar{\mathcal{F}}$, denoted by $S \sim_{s.c.} < g, \beta >$, iff for all $t \geq 0$ and all $x \geq 0$, there holds [10], [11]

$$Prob\{\sup_{0\leq s\leq t}[A \otimes \beta(s) - A^*(s)] > x\} \leq g(x). \quad (7)$$

With the above definitions and their variations, various properties of stochastic network calculus, including stochastic backlog and stochastic delay bounds, have been proved (e.g., see [10], [11], [15]). It has been observed that there are other forms of definitions on stochastic service curves and stochastic arrival curves [10], [11], [15]. The above special forms of definitions have been chosen intentionally to ease the exposition.

III. An Information-Driven Network Calculus

A. Notation on Information

Traditional stochastic network calculus uses the traffic arrival curve and the service curve to count for *cumulative amount* of traffic and service, respectively (i.e., $A(t), S(t) \in R$). If network nodes can perform in-network processing, we need a translation from a flow to the amount of information in the flow in order to model information processing at network nodes. For this purpose, we slightly deviate from the traditional notation of network calculus and use $A(t)$ and $H(A(t))$ to denote the *data* in a flow and the *information* in the flow, respectively, where H is the Shannon entropy function [6]. That is, $A(t)$ denotes a set and H is a function on the set in the rest of the paper. We simply use H to denote information but leave its practical meaning and calculation open to users. This is to avoid different details in entropy estimation in particular applications.

According to [6], the information of a flow $A(t)$, $H(A(t))$, has the following properties:

1) $H(\emptyset) = 0$ where $\emptyset$ denotes the null set, i.e., $H(A(0)) = 0$.
2) $H(A(t))$ is a non-negative, non-decreasing function of time t, since $A(s) \subseteq A(t)$ for $s \leq t$.
3) For different flows $A_i(t), i = 1, \ldots, N$,

$$\sum_{i=1}^{N} H(A_i(t)) \geq H(\sum_{i=1}^{N} A_i(t)), \quad (8)$$

where $\sum_{i=1}^{N} A_i(t)$ means the superposition of flows $A_1(t), \ldots, A_N(t)$.

Definition 3: Information of a set of data sources $A(t) = \{A_1(t), \ldots, A_M(t)\}$, $H(A(t))$, is defined as

$$H(A(t)) \equiv H(\sum_{i=1}^{M} A_i(t)) \quad (9)$$

Definition 4: Redundant information is a measure of the information redundancy in different flows, $A_1(t), \ldots, A_N(t)$. It is denoted as $I(A_1(t); \ldots; A_N(t))$ and is calculated as

$$\sum_{k=1}^{N}(-1)^{k-1} \sum_{\substack{A(t)\subset\{A_1(t),\ldots,A_N(t)\}\\|A(t)|=k}} H(A(t)). \quad (10)$$

Redundant information has the following properties:

1) If $A_1(t), \ldots, A_N(t)$ are independent, then $I(A_1(t); \ldots; A_N(t))$ equals 0.
2) For any $i(= 1, \ldots, N)$, $I(A_1(t); \ldots; A_N(t)) \leq H(A_i(t))$.

Remark 1: Special treatment is required for notation $H(A(s,t))$. By the properties of information, $H(A(t)) = H(A(s)) + H(A(s,t)) - I(A(s); A(s,t))$. Nevertheless, including the above detail will make our later proof unnecessarily lengthy. For this reason, we use

$$\hat{H}(A(s,t)) \equiv H(A(t)) - H(A(s)) = H(A(s,t)) - I(A(s); A(s,t)) \quad (11)$$

to denote the *new* information in the flow during time interval $(s, t]$.

B. Modeling Flow and Service with Respect to Information

When network nodes perform information processing, our model should focus on information instead of packets. For instance, in a sensor network, sensor nodes may be configured to check information redundancy [16], [18], and as such they only care about the amount of information arrivals instead of the amount of packets. This motivates our following definitions:

Definition 5: A flow $A(t)$ is said to have an *information stochastic arrival* curve $\alpha \in \mathcal{F}$ with bounding function $f \in \bar{\mathcal{F}}$, denoted by $A \sim_{i.s.a.} < f, \alpha >$, iff for all $t \geq 0$ and all $x \geq 0$, there holds

$$Prob\{\sup_{0 \leq s \leq t} \sup_{0 \leq u \leq s} [\hat{H}(A(u,s)) - \alpha(s-u)] > x\} \leq f(x). \tag{12}$$

Definition 6: A server is said to provide a flow $A(t)$ with an *information stochastic service* curve $\beta \in \mathcal{F}$ with bounding function $g \in \bar{\mathcal{F}}$, denoted by $S \sim_{i.s.s.} < g, \beta >$, iff for all $t \geq 0$ and all $x \geq 0$, there holds

$$Prob\{\sup_{0 \leq s \leq t} [H(A(s)) \otimes [\beta]^x(s) - H(A^*(s))] > x\} \leq g(x) \tag{13}$$

where $[\beta]^x(t) \equiv \max\{\beta(t), x\}$.

C. Performance Measures

The following definitions are used for information guarantee analysis:

Definition 7: The *information delay* of input flow $A(t)$ in a system at time t is defined as:

$$D(t) = \inf\{\tau \geq 0 : H(A(t)) \leq H(A^*(t+\tau))\}, \tag{14}$$

where $A^*(t)$ is the output flow.

Definition 8: The *information backlog* at time t in a system is defined as:

$$B(t) = H(A(t)) - H(A^*(t)), \tag{15}$$

where $A(t)$ and $A^*(t)$ are input flow and output flow, respectively.

Definition 9: The *information backlog within delay bound* $\tau(\leq D(t))$ at time t is defined as:

$$\hat{B}(t,\tau) = H(A(t)) - H(A^*(t+\tau)). \tag{16}$$

IV. Building Blocks: Basic Properties of Information-Driven Network Calculus

We need to investigate the basic properties of information-driven network calculus. These properties serve as essential building blocks for performance analysis of information-driven networks. To make the paper easy to follow, we move all proofs except the proofs of Theorem 5 (information fusion) and Theorem 7 (Information diffusion) into Appendix.

Lemma 1: [10], [11] For any random variables X and Y, and $\forall x \geq 0$, if $\bar{F}_X(x) \leq f(x)$ and $\bar{F}_Y(x) \leq g(x)$, where $f, g \in \bar{\mathcal{F}}$, then

$$Prob\{X + Y > x\} \leq (f \otimes g)(x). \tag{17}$$

Lemma 2: For any random variables X and Y, and $\forall x \geq 0$, if $\bar{F}_X(x) \leq f(x)$ and $F_Y(x) \leq g(x)$, where $f \in \bar{\mathcal{F}}, g \in \mathcal{F}$, then

$$Prob\{X - Y \geq x\} \leq (f \odot g)(x). \tag{18}$$

Theorem 1: **(Concatenation)** Consider a flow $A(t)$ passing through a network of N nodes in tandem. If each node provides service $S^i \sim_{i.s.s.} < g^i, \beta^i >, i = 1, 2, \ldots, N$, then the network guarantees to the flow a service $S \sim_{i.s.s.} < g, \beta >$ with

$$\beta(t) = \beta^1 \otimes \ldots \otimes \beta^N(t)$$
$$g(x) = g^1 \otimes \ldots \otimes g^N(x).$$

Theorem 1 indicates that the service provided by an end-to-end path follows an *i.s.s.* curve if the nodes along the path follow *i.s.s.* curves.

Theorem 2: **(Output)** Consider a node with input flow A. If $A \sim_{i.s.a.} < f, \alpha >$, and the node provides the flow with service $S \sim_{i.s.s.} < g, \beta >$, then the output $A^* \sim_{i.s.a.} < f \otimes g, \alpha \oslash \beta >$.

Theorem 2 means that the input flow and the output flow both follow *i.s.a.* curves, if the service node follows an *i.s.s.* curve.

Theorem 3: **(Service guarantee)** If the input flow has $A \sim_{i.s.a.} < f, \alpha >$, and the network node provides the flow with service $S \sim_{i.s.s} < g, \beta >$, then

1) The information backlog $B(t)$ of the flow at time t satisfies: for all $t \geq 0$ and all $x \geq 0$,

$$Prob\{B(t) > x\} \leq f \otimes g(x - \alpha \oslash \beta(0)). \tag{19}$$

2) The information delay $D(t)$ of the flow at time t satisfies: for all $t \geq 0$ and all $x \geq 0$,

$$Prob\{D(t) > h(\alpha^x, [\beta]^x)\} \leq f \otimes g(x), \tag{20}$$

where $\alpha^x(t) \equiv \alpha(t) + x$, $[\beta]^x(t) \equiv \max\{\beta(t), x\}$, and $h(\alpha, \beta)$ is the maximum horizontal distance between functions α and β and is defined as

$$h(\alpha, \beta) = \sup_{s \geq 0}\{\inf\{\tau \geq 0 : \alpha(s) \leq \beta(s+\tau)\}\}.$$

3) The information backlog within delay bound $\tau(\leq D(t))$ of the flow at time t satisfies: for all $t \geq 0$ and all $x \geq 0$,

$$Prob\{\hat{B}(t,\tau) > x\} \leq f \otimes g(x + \inf_{v \geq 0}[\beta(v) - \alpha(v - \tau)]). \tag{21}$$

Note that $\alpha(t) = 0$ for $t \leq 0$.

Theorem 3 can be used to calculate the service guarantee provided by a network node in terms of information backlog, information delay, and information backlog within a delay bound.

Theorem 4: **(Service reduction with information impairment)** Consider a network node providing a flow with service $S \sim_{i.s.s.} < g, \beta >$. If the node is interfered with an impairment process $\hat{\mathcal{I}}$ with information stochastic arrival curve $\hat{\mathcal{I}} \sim_{i.s.a.} < f, \alpha >$, then the network node guarantees to the flow a service $S \sim_{i.s.s.} < g \otimes f, \beta - \alpha >$.

Theorem 4 is very useful in some types of network like wireless networks where transmission interference may impact service (note that service includes information processing as well as transmission). We can use this theorem to adjust the service rate of an impacted node.

Theorem 5: **(Information fusion)** Consider two flows $A_1(t)$ and $A_2(t)$. Let $A(t)$ denote the aggregate flow, i.e., $A(t) = A_1(t) + A_2(t)$. If both flows $A_i \sim_{i.s.a.} < f_i, \alpha_i >$

, $i = 1, 2$, and $I(A_1; A_2)$ is lower-bounded by a curve $\gamma \in \mathcal{F}$ with bounding function $\theta \in \mathcal{F}$, such that for all $t \geq 0$ and all $x \geq 0$, there holds

$$Prob\{\inf_{0\leq s\leq t}\inf_{0\leq u\leq s}[I(A_1; A_2)(u, s)) - \gamma(s-u)] \leq x\} \leq \theta(x), \tag{22}$$

then $A \sim_{i.s.a.} < f, \alpha >$, where $f(x) = (f_1 \otimes f_2 \odot \theta)(x)$, and $\alpha(t) = \alpha_1(t) + \alpha_2(t) - \gamma(t)$.

Proof. Based on the properties of information and redundant information, we have

$$\begin{aligned}
&\sup_{0\leq u\leq s}[\hat{H}(A(u, s)) - (\alpha_1(s-u) + \alpha_2(s-u) - \gamma(s-u))] \\
&= \sup_{0\leq u\leq s}[\hat{H}(A_1(u, s)) + \hat{H}(A_2(u, s)) - I(A_1; A_2)(u, s) - \\
&\quad (\alpha_1(s-u) + \alpha_2(s-u) - \gamma(s-u))] \\
&\leq \sup_{0\leq u\leq s}[\hat{H}(A_1(u, s)) - \alpha_1(s-u)] + \sup_{0\leq u\leq s}[\hat{H}(A_2(u, s)) - \\
&\quad \alpha_2(s-u)] - \inf_{0\leq u\leq s}[I(A_1; A_2)(u, s) - \gamma(s-u)]
\end{aligned}$$

for any $s \geq 0$, from which, we further get

$$\begin{aligned}
&\sup_{0\leq s\leq t}\sup_{0\leq u\leq s}[\hat{H}(A(u, s)) - (\alpha_1(s-u) + \alpha_2(s-u) \\
&\quad - \gamma(s-u))] \\
&\leq \sup_{0\leq s\leq t}\sup_{0\leq u\leq s}[\hat{H}(A_1(u, s)) - \alpha_1(s-u)] + \\
&\quad \sup_{0\leq s\leq t}\sup_{0\leq u\leq s}[\hat{H}(A_2(u, s)) - \alpha_2(s-u)] \\
&\quad - \inf_{0\leq s\leq t}\inf_{0\leq u\leq s}[I(A_1; A_2)(u, s) - \gamma(s-u)].
\end{aligned}$$

From the above inequality, the theorem is proved with the definition of *i.s.a.* curve, the fact that $I(A_1; A_2)$ is low-bounded by (22), Lemma 1, and Lemma 2. ■

Theorem 5 means that if two flows follow *i.s.a.* curves, their information fusion (i.e., flow aggregation with information redundancy removed) also follows an *i.s.a.* curve. A **special case** of Theorem 5 is when $A_1(t)$ and $A_2(t)$ are (information) independent. In this case, it is easy to verify that the aggregate flow A has $_{i.s.a.} < (f_1 \otimes f_2)(x), \alpha_1(t) + \alpha_2(t) >$.

We next study information diffusion, which turns out to be harder. When a node performs information diffusion, we model this node as a weighted information splitter. For ease of presentation, we adopt the fluid model in which information could be split in infinitesimal amounts. Clearly, this is the ideal case. For the non-ideal cases, the analysis can be extended by taking into account the information "packetization."

Definition 10: A weighted information splitter is a scheduler that splits an input flow $A(t)$ into multiple *information exclusive* sub-flows $A_1(t), \ldots, A_N(t)$, with each assigned a weight w_i and served by an information processing node $S^i (i = 1, \ldots, N)$, respectively. At any time instant t, the sub-flow assigned to S^i satisfies $H(A_i(t)) = \frac{w_i}{\sum_{j=1}^{N} w_j} H(A(t))$.

For weighted information splitter, the following property can be easily verified:

Theorem 6: Consider a flow $A(t)$ passing through a weighted information splitter with weight parameters

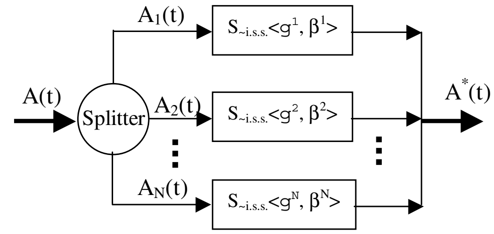

Fig. 2. An information splitter

$w_1, \ldots, w_N$. Denote $\phi_i \equiv \frac{w_i}{\sum_{j=1}^{N} w_j}$ for each sub-flow $A_i(t)$. If the flow has an information stochastic arrival curve $A \sim_{i.s.a.} < f, \alpha >$, then for each sub-flow $A_i(t)$, it also has an information stochastic arrival curve $A_i \sim_{i.s.a.} < f_i, \alpha_i >$, where $\alpha_i(t) = \phi_i \alpha(t)$ and $f_i(x) = f(x/\phi_i)$.

We now consider diffused information passing through a network as shown in Figure 2. In this network, there are N nodes in parallel, and each node provides service $S^i \sim_{i.s.s.} < g^i, \beta^i > (i = 1, 2, \ldots, N)$ to its input. Here, each node may be viewed as the concatenated service effect of an end-to-end path. For such networks, early results in this paper can be used together with Theorem 6 to obtain information delay and backlog at each node as well as the information output from each node. In addition, the network delay, backlog and output can be further derived. With some moderate assumptions as in the following theorem, we can even characterize the service of the whole network to the information flow.

Theorem 7: **(Information diffusion)** Consider a flow $A(t)$ passing through a weighted information splitter with weight parameters and then a network of N nodes in parallel. Also assume that each node provides service $S^i \sim_{i.s.s.} < g^i, \beta^i >$ $(i = 1, 2, \ldots, N)$, where $\beta^i(t) = b^i \cdot t$, and the weight of the sub-flow to S^i is set to $w_i = b^i$. The whole system guarantees to the flow a service $S \sim_{i.s.s.} < g, \beta >$ with

$$\begin{aligned}
\beta(t) &= (\beta^1 + \ldots + \beta^N)(t), \\
g(x) &= g^1 \otimes \ldots \otimes g^N(x).
\end{aligned}$$

Proof. We only prove the case of two parallel nodes, with which, the result can be easily extended to the case of N parallel nodes. Because sub-flows are information exclusive, we have $H(A(t)) = H(A_1(t)) + H(A_2(t))$. Let $\phi_i = \frac{b^i}{b^1+b^2}$, $i = 1, 2$. Because the weight assigned to sub-flow A_i in the weighted information splitter is $w_i = b^i$, we have for any time $s \geq 0$, $H(A_1(s)) = \phi_1 H(A(s))$ and $H(A_2(s)) = \phi_2 H(A(s))$, where $\phi_1 + \phi_2 = 1$.

Note that, for any $s \geq 0$, if $\beta(s) < x$, then $[H(A(s)) \otimes [\beta]^x(s) - H(A^*(s))] \leq [\beta]^x(s) - H(A^*(s)) < x$, which gives no impact in calculating $Prob\{\sup_{0\leq s\leq t}[H(A(s)) \otimes [\beta]^x(s) - H(A^*(s))] > x\}$. In the following, we only consider the case that $\beta(s) \geq x$.

$$
\begin{aligned}
& H(A(s)) \otimes [\beta]^x(s) - H(A^*(s)) \\
= \; & H(A(s)) \otimes [\beta]^x(s) \\
& -H(A_1(s)) \otimes [\beta^1]^x(s) - H(A_2(s)) \otimes [\beta^2]^x(s) \\
& +[H(A_1(s)) \otimes [\beta^1]^x(s) - H(A_1^*(s))] \\
& +[H(A_2(s)) \otimes [\beta^2]^x(s) - H(A_2^*(s))]. \qquad (23)
\end{aligned}
$$

Looking at the first three items, with $\beta(s) \geq x$ and the fact that $\beta^i(s) \leq [\beta^i]^x(s)$, we have

$$
\begin{aligned}
& H(A(s)) \otimes [\beta]^x(s) \\
& -H(A_1(s)) \otimes [\beta^1]^x(s) - H(A_2(s)) \otimes [\beta^2]^x(s) \\
\leq \; & H(A(s)) \otimes \beta(s) - H(A_1(s)) \otimes \beta^1(s) \qquad (24) \\
& -H(A_2(s)) \otimes \beta^2(s) \\
= \; & H(A(s)) \otimes \beta(s) - (\phi_1 H(A(s))) \otimes (\phi_1 \beta(s)) \\
& -(\phi_2 H(A(s))) \otimes (\phi_2 \beta(s)) \\
= \; & H(A(s)) \otimes \beta(s) - \phi_1 \left[H(A(s)) \otimes \beta(s)\right] \\
& -\phi_2 \left[H(A(s)) \otimes \beta(s)\right] \\
= \; & 0
\end{aligned}
$$

Hence, we have

$$
\begin{aligned}
& H(A(s)) \otimes [\beta]^x(s) - H(A^*(s)) \\
\leq \; & [H(A_1(s)) \otimes [\beta^1]^x(s) - H(A_1^*(s))] \\
& +[H(A_2(s)) \otimes [\beta^2]^x(s) - H(A_2^*(s))]. \qquad (25)
\end{aligned}
$$

With simple manipulation, the theorem is proved with the above inequality, the definition of *i.s.s* curve, and Lemma 1.

Remark 2: The above basic properties of information network calculus are strong enough to model most practical information processing systems. For instance, we can use Theorem 5 to handle information sources which exhibit *spatial* and/or *temporal* correlations. We can also model distributed information processing, no matter whether the service or traffic arrivals are correlated or not, because we do not make any assumption on independence.

Remark 3: The information network calculus evolves from and looks similar to traffic-based stochastic network calculus [10], [11], [15]. However, we want to stress that there is significant difference between the information network calculus proposed in the paper and the current development of stochastic network calculus, in that their targeted networks are different. The former is for information-driven networks where the key concern is about the quality of information delivery; the latter and network calculus in general are for networks where traffic is the focus. Due to this fundamental difference, special care has to be taken in developing the calculus for information-driven networks. For example, information dependence and information redundancy are unique concepts for information-driven networks. While two traffic flows may be independent, they can carry the same or highly correlated information as discussed in Example 1.

Remark 4: The theorems in this section hold under the implicit assumption that there is no *information loss* during transmission. This assumption does not contradict Example 1 where packets may be lost, because in many situations packet loss does not necessarily mean information loss. Problems in an information-lossy system require special care and are left as our future work. Briefly, we may need a "clipper" like component [14] to handle the information loss.

V. Revisiting Example 1

We use the example in Section I to illustrate how to use our information calculus for performance evaluation.

Temporal correlation. Assume that the information of each sensor is collected by periodically sampling a stationary Gaussian stochastic process. Concretizing the information $H(\cdot)$ of the flow A_i generated by source i to the Shannon entropy function, it yields for discrete time t [6]:

$$H(A_i(t)) = \alpha_i(t) = \frac{1}{2}\log(2\pi e)^t |\mathcal{C}_i^{(t)}|, \;\; t = 1, 2, ... \qquad (26)$$

where $\mathcal{C}_i^{(t)}$ is the $t \times t$ covariance matrix for the flow of source i and is specified by the temporal covariance function $\Gamma_i(\tau)$, i.e., the matrix element $\mathcal{C}_i^{(t)}(j,k) = \Gamma_i(k-j)$, where $1 \leq j,k \leq t$. Here, we adopt the typical exponential covariance function [20]:

$$\Gamma_i(\tau) = \sigma_i^2 e^{-|\tau|/\eta_i}, \;\; \tau = 0, \pm 1, \pm 2, ... \qquad (27)$$

where σ_i^2 is the variance of flow A_i and η_i is a constant. Assume that each source generates data at a constant interval δ. By applying Equation (27) in (26), the stochastic arrival curve of flow A_i can be specified in the continuous time t as $A_i \sim_{i.s.a.} < 0, \alpha_i >$, where

$$
\alpha_i(t) = \qquad (28)
$$
$$
\begin{cases} \frac{t}{2\delta}\log(2\pi e \sigma_i^2), 0 \leq t \leq \delta \\ \frac{t}{2\delta}\log(2\pi e \sigma_i^2(1 - e^{-2/\eta_i})) - \frac{1}{2}\log(1 - e^{-2/\eta_i}), t > \delta \end{cases}
$$

Spatial correlation. We model the information redundancy of the sources using a spatial correlation model similar to that in [20]. Specifically, we assume that for the 4 sources,

$$H(\sum_{i=1}^{4} A_i) = \sum_{i=1}^{4} \epsilon_i H(A_i), \;\; 0 \leq \epsilon_i \leq 1 \qquad (29)$$

where ϵ_i are constants and depend on the sources' locations and the adopted spatial model [20]. We set all $\epsilon_i = 0.5$ as an example.

Information processing. Assume that each of the two relay nodes provides information service $\sim_{i.s.s.} < e^{-x}, r(t-d) >$, where r is the average information service rate and d is the per-hop delay. Assume that the two nodes are subject to correlated impairment $\sim_{i.s.a.} < 4e^{-x/4}, r(t-d)/5 >$. We emphasize that **information fusion** is performed with (29) when multiple correlated information sources arrive at the relay nodes.

Results. We select the rate parameter $r = 4$ kbps, per-hop delay $d = 7.5$ ms. For every information source, we set $\delta = 100$ ms and $\eta_i = 100$. With reference to Equation (28), σ_i is set such that the long-term information rate of a source

$\approx$ 2.33 kbps. Considering the spatial correlation, the total long-term information arrival rate of the 4 sources amounts to 4.66 kbps. With the theorems in Section III, we obtain the following results:

1) The stochastically achievable information service of the system follows $\sim_{i.s.s.} < 5e^{-x/5}, 6.4(t-0.015)(kbps) >$.
2) The total information backlog in the system, $B(t)$ satisfies: for all $t \geq 0$ and all $x \geq 0$, $Prob\{B(t) > x\} \leq 5e^{-(x-0.096)/5}$.
3) For any information source, the information delay $D(t)$ at time t satisfies: for all $t \geq 0$ and all $x \geq 0$, $Prob\{D(t) > x\} \leq 5e^{-1.28(x-0.015)}$.

VI. Related Work

To the best of our knowledge, there are currently no analytical tools available for systematic performance study of information-driven networks. The framework proposed in this paper is related to *network calculus*, particularly its stochastic branch: *stochastic network calculus*. Since its introduction in early 1990s [7], network calculus has attracted a lot of research attention and evolved along two tracks – deterministic and stochastic. Excellent books summarizing results for deterministic network calculus are available (e.g. [4], [14]). For stochastic network calculus, its research can also be tracked back to early 1990s (e.g. [13], [21]). However, due to some difficulties specific to stochastic networks [10], [11], [15], it is only in recent years when critical network calculus properties such as concatenation property [5], [10], [11] and independent case analysis [10], [11] have been proved for stochastic network calculus. The relevance of the present paper to stochastic network calculus lies in the analogy between the various models and properties defined or derived in this paper for information-driven networks, and the corresponding models and properties under stochastic network calculus. Nevertheless, they are different according to Remark 3.

In information-driven networks, in-network information processing is likely performed. From this in-network processing viewpoint, the present paper is related to [8] and [19]. In [8], scaling functions are used to model the relationship between the input traffic and the output traffic of a network element that processes the traffic. Based on the proposed scaling server model, [8] extends the deterministic network calculus by considering data scaling in networks with in-network data processing. In [19], how the scaling elements can be shifted across multiplexers in sensor networks is studied, which enables worst-case analysis of traffic delay and backlog in such networks. Note that in [8], [19], information processing only applies to intra-flow data, leaving inter-flow processing un-considered. However, in our work, both intra-flow processing and inter-flow processing are considered. In addition, the essential focus of [8] and [19] is on traffic, while our focus is on information carried by traffic.

In [1], the problem of *network information flow* is introduced. The focus of [1] is on a special type of in-network processing which is called *network coding*. With network coding, information is diffused through the network from the sources to the destination(s) and sources of flows may be jointly coded to achieve optimality in addressing the network information flow problem. An excellent introduction to network coding theory is available [22]. The present paper is related to [1] and network coding literature in that they all take *information* as the central point of study.

Note that with network coding, sources may be coded jointly. In such cases, focusing on traffic in the analysis may not be applicable, since the output flow is not a simple scaling of the input flow or the aggregate of input flows. For example, suppose flows f_1 and f_2 are two bit streams. Applying exclusive-OR to the corresponding bits in them results in a *new* flow $f_1 \oplus f_2$. In the current network calculus literature including [8] and [19], traffic amount and (traffic) service amount are the concern. In the example, for traffic, $A_{f_1 \oplus f_2}(t) = A_{f_1}(t) = A_{f_2}(t)$. For service to each individual flow or superposition of these two flows, however, the current network calculus approach provides no answer, since no corresponding output of f_1 or f_2 is found out of the exclusive-OR operation.

We would like to highlight that network coding and other in-network processing techniques significantly complicate network performance analysis both in terms of traffic service guarantees and in terms of information service guarantees. While a lot of network calculus results are available for potential use in analyzing *traffic* service guarantees in such networks (e.g., [17]), no previous work has been found for analyzing *information* service guarantees in these networks. We believe the proposed calculus makes a critical step and sheds light on further development to address this challenge.

VII. Conclusions and Open Research Challenges

As network architectures are re-evaluated and extended to set aside traditional layering principle in challenged environments, in-network processing of data is becoming much more important. QoS guarantee in this type of networks should be measured with respect to quality of information instead of just data throughput or bounded (end-to-end) packet delay. Although substantial research has been done in information processing for specific applications, a systematic analytical framework for performance modeling and evaluation of information-driven networks remains blank. This paper tries to fill this gap.

This paper focuses mainly on the development of a new calculus, which lays a foundation for analytical performance evaluation of information-driven networks. We expect that it will stimulate follow-on research and push system design toward more fundamental principles. Along the line, many research challenges demand further investigation, including, for example, (1) the stochastically achievable information rates with a generic network topology, (2) the various optimization problems in information fusion and diffusion, (3) the performance bounds if information lossy models are introduced.

Acknowledgment

This work was supported by a fellowship from the Centre for Quantifiable Quality of Service in Communication Systems at the Norwegian University of Science and Technology, and by the Discovery Grant of Natural Sciences and Engineering Research Council of Canada.

References

[1] R. Ahlswede, N. Cai, S. Li, and R. Yeung. Network information flow. *IEEE Trans. Information Theory*, 46(4):1204–1216, 2000.

[2] R. Ahmed and R. Boutaba. Distributed pattern matching: A key to flexible and efficient p2p search. *IEEE Journal on Selected Areas in Communications*, 25(1):73–83, January 2007.

[3] G. Armitage. *Quality of Service in IP Networks*. Macmillian TP, 2000.

[4] C.-S. Chang. *Performance Guarantees in Communication Networks*. Springer-Verlag, 2000.

[5] F. Ciucu, A. Burchard, and J. Liebeherr. A network service curve approach for the stochastic analysis of networks. *IEEE Trans. Information Theory*, 52(6):2300–2312, June 2006.

[6] T. M. Cover and J. A. Thomas. *Elements of Information Theory, Second Edition*. John Wiley & Sons, 2006.

[7] R. L. Cruz. A calculus for network delay, part I: network elements in isolation. *IEEE Trans. Information Theory*, 37(1):114–131, Jan. 1991.

[8] M. Fidler and J. B. Schmitt. On the way to a distributed systems calculus: An end-to-end network calculus with data scaling. In *Proc. ACM SIGMETRICS/Performance 2006*, pages 287–298, 2006.

[9] C. Intanagonwiwat, R. Govindan, and D. Estrin. Directed diffusion: A scalable and robust communication paradigm for sensor networks. In *Proceedings of ACM MobiCOM '00*, pages 56–67, Boston, August 2000.

[10] Y. Jiang. A basic stochastic network calculus. In *Proceedings of ACM Sigcomm 06*, pages 123–134, Pisa, Italy, September 2006.

[11] Y. Jiang and Y. Liu. *Stochastic Network Calculus*. Springer, 2008.

[12] D. Kundur, Z. Liu, M. Merabti, and H. Yu. Advances in peer-to-peer content search. In *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, pages 404–407, Beijing, July 2007.

[13] J. Kurose. On computing per-session performance bounds in high-speed multi-hop computer networks. In *ACM SIGMETRICS'92*, 1992.

[14] J. Le Boudec and P. Thiran. *Network Calculus: A Theory of Deterministic Queuing Systems for the Internet*. Springer-Verlag, 2001.

[15] C. Li, A. Burchard, and J. Liebeherr. A network calculus with effective bandwidth. *IEEE/ACM Transactions on Networking*, 15(6):1142–1453, December 2007.

[16] C. Liu, K. Wu, and J. Pei. An energy efficient data collection framework for wireless sensor networks by exploiting spatiotemporal correlation. *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, 18(7):1010–1023, July 2007.

[17] A. Mahmino, J. Lecan, and C. Fraboul. Guaranteed packet delays with network coding. In *First IEEE International Workshop on Wireless Network Coding*, pages 1–6, San Francisco, June 2008.

[18] S. Pattem, B. Krishnamachari, and R. Govindan. The impact of spatial correlation on routing with compression in wireless sensor networks. *ACM Transactions on Sensor Networks (TOSN)*, 4(4), August 2008.

[19] J. B. Schmitt, F. A. Zdarsky, and L. Thiele. A comprehensive worst-case calculus for wireless sensor networks with in-network processing. In *Proceedings of IEEE RTSS 2007*, pages 193–202, Tucson, Arizona, December 2007.

[20] M. Vuran, O. B. Akan, and I. Akyildiz. Spatio-temporal correlation: Theory and applications for wireless sensor networks. *Computer Networks*, 45(3):245–259, June 2004.

[21] O. Yaron and M. Sidi. Performance and stability of communication network via robust exponential bounds. *IEEE/ACM Trans. Networking*, 1(3):372–385, June 1993.

[22] R. Yeung, S. Li, N. Cai, and Z. Zhang. *Network Coding Theory*. now Publishers, 2006.

Appendix

Proof of Lemma 2. For any random variables X and Y, and any $x \geq 0, z \geq 0$, $\{X - Y \geq x\} \cap \{X \leq x + z\} \cap \{Y > z\} = \emptyset$, where $\emptyset$ denotes the null set. We thus have

$$\{X - Y \geq x\} \subseteq \{X > x + z\} \cup \{Y \leq z\},$$

which means

$$Prob\{X - Y \geq x\} \leq Prob\{X > x + z\} + Prob\{Y \leq z\}.$$

Since the above inequality holds for all $z \geq 0$, we get

$$Prob\{X - Y \geq x\} \leq \inf_{z \geq 0}[Prob\{X > x + z\} + Prob\{Y \leq z\}],$$

with which and $\bar{F}_X(x) \leq f(x)$ and $F_Y(x) \leq g(x)$, where $f \in \bar{\mathcal{F}}, g \in \mathcal{F}$, the result is proved.■

Proof of Theorem 1. We only prove the two node case, because the same result can be extended to the N-node case. We use $A^i(t)$ and $A^{i*}(t)$ to denote the input flow and the output flow of node i, respectively. Note that $A^{1*}(t) = A^2(t)$. For any $s \geq 0$ and $x \geq 0$, we have

$$\begin{aligned}
& H(A^1(s)) \otimes [\beta^1 \otimes \beta^2]^x(s) - H(A^{2*}(s)) \\
\leq & H(A^1(s)) \otimes ([\beta^1]^x \otimes [\beta^2]^x)(s) - H(A^{2*}(s)) \\
= & \inf_{0 \leq u \leq s}[H(A^1(u)) \otimes [\beta^1]^x(u) + [\beta^2]^x(s-u) - H(A^{1*}(u)) \\
& + H(A^2(u))] - H(A^{2*}(s)) \\
\leq & \sup_{0 \leq u \leq s}[H(A^1(u)) \otimes [\beta^1]^x(u) - H(A^{1*}(u))] \\
& + \inf_{0 \leq u \leq s}[H(A^2(u)) + [\beta^2]^x(s-u)] - H(A^{2*}(s)) \\
= & \sup_{0 \leq u \leq s}[H(A^1(u)) \otimes [\beta^1]^x(u) - H(A^{1*}(u))] \\
& + H(A^2(s)) \otimes [\beta^2]^x(s) - H(A^{2*}(s)).
\end{aligned}$$

We thus have for any $t \geq 0$,

$$\begin{aligned}
& \sup_{0 \leq s \leq t}[H(A^1(s)) \otimes [\beta^1 \otimes \beta^2]^x(s) - H(A^{2*}(s))] \\
\leq & \sup_{0 \leq u \leq t}[H(A^1(u)) \otimes [\beta^1]^x(u) - H(A^{1*}(u))] \\
& + \sup_{0 \leq s \leq t}[H(A^2(s)) \otimes [\beta^2]^x(s) - H(A^*(s))].
\end{aligned}$$

The theorem is proved with the above inequality, the definition of information stochastic service curve, and Lemma 1. ■

Proof of Theorem 2. Based on the properties of information, for any $0 \leq s \leq t$, we have

$$\begin{aligned}
& \hat{H}(A^*(s,t)) = H(A^*(t)) - H(A^*(s)) \\
\leq & H(A(t)) - H(A^*(s)) \\
= & H(A(t)) - H(A(s)) \otimes [\beta]^x(s) + H(A(s)) \otimes [\beta]^x(s) \\
& - H(A^*(s)) \\
= & \sup_{0 \leq u \leq s}[H(A(t)) - H(A(u)) - [\beta]^x(s-u)] \\
& + [H(A(s)) \otimes [\beta]^x(s) - H(A^*(s))] \\
= & \sup_{0 \leq u \leq s}[\hat{H}(A(u,t)) - \alpha(t-u) + \alpha(t-u) - [\beta]^x(s-u)] \\
& + [H(A(s)) \otimes [\beta]^x(s) - H(A^*(s))] \\
\leq & \sup_{0 \leq u \leq t}[\hat{H}(A(u,t)) - \alpha(t-u)] + \alpha \oslash \beta(t-s) \\
& + [H(A(s)) \otimes [\beta]^x(s) - H(A^*(s))]
\end{aligned}$$

We thus have

$$\begin{aligned}
& \sup_{0 \leq s \leq t}[\hat{H}(A^*(s,t)) - \alpha \oslash \beta(t-s)] \\
\leq & \sup_{0 \leq u \leq t}[\hat{H}(A(u,t)) - \alpha(t-u)] \\
& + \sup_{0 \leq s \leq t}[H(A(s)) \otimes [\beta]^x(s) - H(A^*(s))]
\end{aligned}$$

from which together with simple manipulation, we further get

$$\sup_{0\le s\le t}\sup_{0\le u\le s}[\hat{H}(A^*(u,s))-\alpha\oslash\beta(s-u)]$$
$$\le \sup_{0\le s\le t}\sup_{0\le u\le s}[\hat{H}(A(u,s))-\alpha(s-u)]$$
$$+\sup_{0\le s\le t}[H(A(s))\otimes[\beta]^x(s)-H(A^*(s))].$$

From the above inequality, the theorem is proved with the definition of information stochastic arrival curve, the definition of information stochastic service curve, and Lemma 1. ■

Proof of Theorem 3. 1) For the information backlog $B(t)$, we have for any $t, x \ge 0$,

$$B(t)=H(A(t))-H(A^*(t))$$
$$=H(A(t))-H(A(t))\otimes[\beta]^x(t)+H(A(t))\otimes[\beta]^x(t)-H(A^*(t))$$
$$=\sup_{0\le s\le t}[\hat{H}(A(s,t))-\alpha(t-s)+\alpha(t-s)-[\beta]^x(t-s)]$$
$$+[H(A(t))\otimes[\beta]^x(t)-H(A^*(t))]$$
$$\le \sup_{0\le s\le t}[\hat{H}(A(s,t))-\alpha(t-s)]+\alpha\oslash\beta(0)$$
$$+[H(A(t))\otimes[\beta]^x(t)-H(A^*(t))]$$
$$\le \sup_{0\le s\le t}[\hat{H}(A(s,t))-\alpha(t-s)]+\alpha\oslash\beta(0)$$
$$+\sup_{0\le s\le t}[H(A(s))\otimes[\beta]^x(s)-H(A^*(s))]$$

The result is proved with the above inequality, the definition of information stochastic arrival curve, the definition of information stochastic service curve, and Lemma 1.

2) For the information delay $D(t)$, we have from the definition, for any $y\ge 0$, $\{D(t)>y\}\subset\{H(A(t))>H(A^*(t+y))\}$ and hence $Prob\{D(t)>y\}\le Prob\{H(A(t))>H(A^*(t+y))\}$. We also have:

$$H(A(t))-H(A^*(t+y))$$
$$=H(A(t))-H(A(t+y))\otimes[\beta]^x(t+y)+H(A(t+y))$$
$$\otimes[\beta]^x(t+y)+\alpha(t-s)-\alpha(t-s)-H(A^*(t+y))$$
$$\le \sup_{0\le s\le t}[\hat{H}(A(s,t))-\alpha(t-s)]$$
$$+H(A(t+y))\otimes[\beta]^x(t+y)-H(A^*(t+y))$$
$$+\sup_{0\le s\le t+y}[\alpha(t-s)-[\beta]^x(t-s+y)]$$

By replacing y with $h(\alpha+x,[\beta]^x)$ in above, where $h(\alpha+x,[\beta]^x)=\sup_{s>0}\{\inf\{\tau\ge 0:\alpha(s)+x\le[\beta]^x(s+\tau)\}\}$ is the maximum horizontal distance between functions $\alpha(t)+x$ and $[\beta]^x(t)$, which implies $\alpha(t)+x\le[\beta]^x(t+h(\alpha+x,[\beta]^x))$, we obtain

$$H(A(t))-H(A^*(t+h(\alpha+x,[\beta]^x)))$$
$$\le \sup_{0\le s\le t}[\hat{H}(A(s,t))-\alpha(t-s)]$$
$$+H(A(t+h(\alpha+x,[\beta]^x)))\otimes[\beta]^x(t+h(\alpha+x,[\beta]^x))$$
$$-H(A^*(t+h(\alpha+x,[\beta]^x)))-x$$
$$\le \sup_{0\le s\le t}[\hat{H}(A(s,t))-\alpha(t-s)]+$$
$$\sup_{0\le s\le t}[H(A(s+h(\alpha+x,[\beta]^x)))\otimes[\beta]^x(s+h(\alpha+x,[\beta]^x))$$
$$-H(A^*(s+h(\alpha+x,[\beta]^x)))]-x$$

Based on the above inequality, the definition of information stochastic arrival curve, the definition of information stochastic service curve, and Lemma 1, we have $Prob\{D(t)\}>h(\alpha+x,[\beta]^x)\}\le f\otimes g(x)$.

3) For the information backlog within delay bound $\tau(\le D(t))$, using the same derivation as in (2), we have

$$\hat{B}(t,\tau)=H(A(t))-H(A^*(t+\tau))$$
$$\le \sup_{0\le s\le t}[\hat{H}(A(s,t))-\alpha(t-s)]$$
$$+H(A(t+\tau))\otimes[\beta]^x(t+\tau)-H(A^*(t+\tau))$$
$$+\sup_{0\le s\le t+\tau}[\alpha(t-s)-[\beta]^x(t-s+\tau)]$$
$$=\sup_{0\le s\le t}[\hat{H}(A(s,t))-\alpha(t-s)]$$
$$+H(A(t+\tau))\otimes[\beta]^x(t+\tau)-H(A^*(t+\tau))$$
$$-\inf_{v\ge 0}[[\beta]^x(v)-\alpha(v-\tau)]$$
$$\le \sup_{0\le s\le t}[\hat{H}(A(s,t))-\alpha(t-s)]$$
$$+\sup_{0\le s\le t}[H(A(s+\tau))\otimes[\beta]^x(s+\tau)-H(A^*(s+\tau))]$$
$$-\inf_{v\ge 0}[\beta(v)-\alpha(v-\tau)]$$

The result is proved with the above inequality, the definition of information stochastic arrival curve, the definition of information stochastic service curve, and Lemma 1.■

Proof of Theorem 4. We could treat the system as if it provides service to the "aggregate" of two flows: the input $A(t)$ and the impairment $\hat{\mathcal{I}}(t)$. Denote the "aggregate" as $\hat{A}(t)$. There holds[1] $H(\hat{A}(t))=H(A(t))+H(\hat{\mathcal{I}}(t))$ and $H(\hat{A}^*(t))=H(A^*(t))+H(\hat{\mathcal{I}}^*(t))$, where $\hat{A}^*(t), A^*(t)$, and $\hat{\mathcal{I}}^*(t)$ are the outputs of $\hat{A}(t), A(t)$, and $\hat{\mathcal{I}}(t)$, respectively. We have for any $s, x\ge 0$,

$$H(A(s))\otimes[\beta-\alpha]^x(s)-H(A^*(s))$$
$$\le \inf_{0\le u\le s}[H(\hat{A}(u))-H(\hat{\mathcal{I}}(u))+[\beta]^x(s-u)-\alpha(s-u)]$$
$$-H(\hat{A}^*(s))+H(\hat{\mathcal{I}}^*(s))$$
$$\le [H(\hat{A}(s))\otimes[\beta]^x(s)-H(\hat{A}^*(s))]-\inf_{0\le u\le s}[H(\hat{\mathcal{I}}(u))+\alpha(s-u)]$$
$$+H(\hat{\mathcal{I}}^*(s))$$
$$\le [H(\hat{A}(s))\otimes[\beta]^x(s)-H(\hat{A}^*(s))]-\inf_{0\le u\le s}[H(\hat{\mathcal{I}}(u))+\alpha(s-u)]$$
$$+H(\hat{\mathcal{I}}(s))$$
$$=[H(\hat{A}(s))\otimes[\beta]^x(s)-H(\hat{A}^*(s))]$$
$$+\sup_{0\le u\le s}[\hat{H}(\hat{\mathcal{I}}(u,s))-\alpha(s-u)]$$

We hence have for any $t\le 0$,

$$\sup_{0\le s\le t}[H(A(s))\otimes[\beta-\alpha]^x(s)-H(A^*(s))]$$
$$\le \sup_{0\le s\le t}[H(\hat{A}(s))\otimes[\beta]^x(s)-H(\hat{A}^*(s))]$$
$$+\sup_{0\le s\le t}\sup_{0\le u\le s}[\hat{H}(\hat{\mathcal{I}}(u,s))-\alpha(s-u)].$$

The theorem is proved because the system provides $\hat{A(t)}$ with service $\sim_{i.s.s.}< g,\beta >$ and the impairment process $\hat{\mathcal{I}}$ follows $\sim_{i.s.a.}< f,\alpha >$.■

[1] For simplicity, we assume that $A(t)$ and $\hat{\mathcal{I}}(t)$ are independent, otherwise the theorem needs a trivial modification to accommodate the redundant information between $A(t)$ and $\hat{\mathcal{I}}(t)$.