This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2707597, IEEE Access



# Congestion Control in Molecular Cyber-Physical Systems

Luca Felicetti, Mauro Femminella, *Member, IEEE,* Gianluca Reali, *Member, IEEE*

***Abstract*—Cyberphysical systems (CPSs) are a new class of engineered systems based on interactions between cyber and physical components, by integrating three main components: communications, control, and computing. When these systems are brought to the nanoscale, some design and implementation issues arise. A high level of complexity is due to the use of biological components in a CPS, such as engineered cells, which may play the role of sensors, actuators, or even controllers. In this paper we study the effectiveness of control solutions implemented through the usage of molecular communications in a biological nanoscale cyber-physical system (BioNanoCPS), where a biological nanomachine plays the role of actuator, that releases drug molecules, and another acts as both sensor and controller. The goal of the proposal is to control the release rate, so that target cells can receive the desired amount of drug in a given time, by limiting potential side effects. Basically, we aim to limit congestion, which can arise when large amounts of molecules are released towards a target. To this aim, we propose a simple congestion detection scheme, and compare different rate control algorithms used to throttle the molecules release rate at the transmitter upon the reception of a feedback signal sent by the receiver. We validate the proposed techniques against delivery efficiency and delivery time of molecules by means of an extensive simulation campaign.**

***Index Terms*—Congestion detection, congestion control, feedback-based rate control, diffusion-based molecular communications.**

## I. Introduction

CYBER-PHYSICAL systems (CPSs) are a new class of engineered systems based on interactions between cyber and physical components, by integrating three main components: communications, control, and computing [1]. CPSs are expected to play a major role in research, design, and implementation of future technologies. When these systems include devices at the nanoscale, some design and implementation issues arise. A high level of complexity is due to the use of biological components, such as engineered cells, which may act as sensors, actuators, or even controllers in living bodies. Such systems take clearly advantage of the research results in the field of molecular communications (MolCom) [2]–[4]. MolCom is a novel communication paradigm proposed for connecting biological nanomachines (specifically bio-nanomachines) over short ranges, within aqueous environments. In synthesis, MolCom systems allow artificial nanomachines, including engineered cells, to communicate through the exchange of molecules, by mimicking the natural cell signaling. In this model, a transmitter nanomachine emits molecules, which act as communication signals, that are received by receiver nanomachines. The reception process happens through chemical reactions between signal molecules (ligand) and compliant receptors present on the receiver surface. Given the bio-compatibility of these systems, they could allow implementing closed loop systems (which make them CPS), which offer the possibility to tightly monitor and control biological processes. These features are expected to give birth to really disruptive applications in the fields of medicine, food production, and so on.

In this paper, we propose the use a biological, nanoscale cyber-physical system (BioNanoCPS) to control the process of drug delivery/tissue engineering [5]–[7] directly *in situ*. We assume that some biological nanomachines (bio-nanomachines) act as actuators, that are able to release drug molecules, and another bio-nanomachine behaves as both sensor and controller. When the controller detects the proximity of a given target (e.g. a tumor), it triggers the actuators for releasing drugs by a command sent by using a MolCom system. However, since the distance between the actuators and and the target is not known in advance, the release rate has to be adapted in order to avoid wasting of drug molecules. Thus, the first tasks of the controller is to monitor the drug absorption rate (sensing function), by absorbing the emitted drugs as if it were the receiver. The second one is to control the drug release rate. The goal of this second function it twofold. It both allows target cells to receive the desired amount of drug in a given time and, at the same time, it avoids to release an amount of drug that cannot be absorbed by the target, which could even produce undesired side effects and drug waste. Also the messages used to control the release rate are sent by using the MolCom system.

In this paper, we focus on the effectiveness of the control solutions. In particular, we aim to limit the phenomenon of congestion due to accumulation of drug molecules, already introduced in [8] and further elaborated in [9]. This phenomenon can arise when a large amount of molecules are released towards a target. We propose both a simple congestion detection scheme (first contribution), which can be implemented in a bio-nanomachine, and four different rate control algorithms (second contribution), all based on the negative feedback concept [10]. We compare these different rate control algorithms used to throttle the molecules release rate at the transmitter upon the reception of a feedback signal, sent by the receiver. Some of them inspire to the control mechanisms implemented in the TCP protocol, mentioned in [11]. Without any loss of generality, we use the communication protocols already defined in [11], which allows focusing only on the effectiveness of control algorithms. We have assessed

Authors were with the Department of Engineering, University of Perugia, CNIT Research Unit, via G. Duranti 93, 06125 Perugia, Italy, email: ing.luca.felicetti@gmail.com, mauro.femminella@unipg.it, gianluca.reali@unipg.it.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2707597, IEEE Access



the performance of the proposed techniques against delivery efficiency and delivery time of molecules. Performance have been evaluated by means of an extensive simulation campaign, implemented by using the BiNS2 simulator [12], [13].

The paper is organized as follows. In section II, we illustrate the related work in the field. Section III presents the system model. In section IV, we illustrate our proposed congestion detection scheme. Section V shows our proposal for control actions, to be implemented in the TX node, and the relevant decision logic, to be implemented in the RX node. The results of the simulation campaign, used to validate the proposed mechanisms and to analyze the trade-off between throughput and communication efficiency, are presented in Section VI. Finally, we draw our conclusions in Section VII.

## II. Background and related works

Major efforts in the area of Molecular Communications are focused on the physical layer issues of various types of MolCom transmission media. Information capacity and physical features (e.g., delay, signal attenuation, amplification, and energy requirements) of MolCom systems are typically studied by using random walk models [14]–[16], random walk models with drift [17], [18], diffusion-reaction-based models [19], [20], active transport models [14], [21], collision-based models [22], and diffusion-based models [23]–[27], which is used in this paper.

Most of results have been obtained by assuming that a sufficiently high density of receptors are deployed on the receiver surface, so that all the available molecules coming in contact with the receiver surface come also in contact with the receptors and are absorbed (absorbing receiver assumption, [28]). A more refined model, which makes use of a finite number of receptors with a specific size, is presented in [29]. Although this model is more realistic, it assumes that each time a ligand molecule comes in contact with a compliant receptor, it is immediately absorbed (absorbing receptor assumption). Hence, the stochastic nature of ligand-receptor binding is not considered. Differently, a finite bonding duration is analyzed in [30], and further investigated in [31], which takes also into account the trafficking time [32], which can be essentially regarded as the molecule reception time. Specifically, it is defined as the time between the two following events. The first is the bond formation between a ligand molecule and a receptor. THe second is the internalization of the ligand-receptor complex. After this internalization, the receiver can expose another free receptor on its surface. A reversible receiver model, which does not explicitly model receptors but rather their macroscopic behavior by means of a reaction constant, is presented in [33].

Finally, the papers [8], [9] analyze the congestion phenomenon in molecular communications, and provide theoretical models to predict its occurrence when a continuous stream of drugs are released towards a target. In this paper, by making use of these results, we develop a detection scheme able to early identify congestion conditions when a flow of molecules is released towards a target with variable rate and unknown communication range.

## III. The system model

The considered communication scenario consists of two fixed bio-nanomachines at a distance $d$: one acting as receiver and control node (RX), and the other as transmitting node (TX). Both TX and RX are modeled as spherical nanomachines, with radius $r_{TX}$ and $r_{RX}$, respectively. This is a simplified model, since it does not explicitly consider the effect of the target (e.g. tumor cells) on the sensing process carried out by the RX. However, an estimation of the presence of additional receivers (i.e. target cells) can be obtained by using the results presented in [34]. The communication happens by using two different types of molecules, which propagate by diffusion, modeled as Brownian motion [35]. The molecules transmitted by the TX node, and representing the signal to be delivered, are labeled as $S$, and those transmitted by the control node RX, to encode the control messages, are labeled as $R$. The Brownian motion is characterized by the diffusion coefficient, given by $D = \frac{K_b T}{6\pi\eta r_c}$. $K_b$ is the Boltzmann constant, $T$ denotes the temperature expressed in Kelvins, $\eta$ is the viscosity of the medium, and $r_c$ is the radius of the considered molecules, i.e. $r_{c,rx}$ for type $R$ molecules and $r_{c,tx}$ for type $S$ molecules.

When a ligand molecule hits a free, compliant receptor with radius $r_r$, a bond is established. The trafficking time is modeled as an exponential random variable [32], with mean $T_{traff}$. A number of receptors ($R_{TX}$ on the TX for type $R$ molecules and $R_{RX}$ on the RX for type $S$ ones) is distributed over the surface of each nanomachine.

Without any loss of generality, we use the protocol defined in [11], where the control signals are encoded in different patterns of bursts, with size $B_{RX}$ of $R$ molecules, by using the OOK modulation, and symbol duration $T_S$. The defined control messages are: START, encoded as 110, CONTROL, encoded as 10, and STOP, encoded as 111. For decoding the control bursts, the TX node adopts a threshold mechanism, with value $\zeta_S$. Motivations about the design choices of the control protocol can be found in [11], whereas the values used in simulations are reported in Table I. Any other connection-oriented protocol can be used to the goal of this paper. The selection of using the one proposed in [11] is due to the fact that, to the best of the authors' knowledge, it is the only one currently proposed in the MolCom literature.

Since in operation the distance between RX and TX is unknown, the TX "probes" the channel with an increasing rate of $S$ molecules in order to determine the most suitable carrier release rate. This process is controlled by the RX, which sends throttling (CONTROL) messages when congestion is detected. Specifically, the TX nodes starts with a burst $B_{TX} = B_{0,TX}$ of $S$ molecules and continues increasing the transmission burst $B_{TX}$ by $B_{0,TX}$ $S$ molecules each time interval $\Delta t$. Thus, the TX increases the release rate with a linear slope, until it receives either a STOP or a CONTROL message from the controller RX. Different reaction strategies to CONTROL messages at the TX node are presented in section V.

At the RX site, connections are successfully established when $\zeta_{RTT}$ $S$ molecules are received by a suitable time and, at the same time, also the round time time (RTT) of the communication path is estimated. The instant when the RTT

is estimated and the connection is set up is denoted $t_{\zeta_{RTT}}$. The estimation of the RTT is equal to the time elapsed since the transmission of the final burst of the START signal from the RX to the time instant $t_{\zeta_{RTT}}$. Since $t_{\zeta_{RTT}}$, the RX starts monitoring the total amount of received $S$ molecules $N(t)$, updated with a period equal to $T_w$. The reception rate of the $S$ molecules is denoted $\lambda_a(t)$, thus $N(t) = \int_0^t \lambda_a(t)$. However, when a molecule hits a compliant receptor already busy in another bond, it cannot establish a bond with it and we say that it is "rejected", as in an already full queuing system [8], [9]. The rate of such rejections is denoted $\lambda_r(t)$. When a rejection happens, the ligand molecule is bounced back as a result of a partially inelastic collision. In [33], an alternative model to reposition a detached molecule that breaks a bond is presented.

The transmission is ended by the RX by sending a STOP message when either it estimates that the target amount of absorbed molecules $\zeta_{stop}$ has been already reached (i.e. $N(t) \geq \zeta_{stop}$) or it will be reached in the next few seconds by issuing a STOP message (i.e. $N_{stop}(t) \geq \zeta_{stop}$).

## IV. Detection of congestion

Intuitively, the RX can estimate the emergence of congestion if the estimated absorption rate deviates from its expected behavior. Since the RX knows the law used by the TX to transmit the signal (i.e. a linear increase by $B_{0,TX}$ of the burst size $B_{TX}$ each $\Delta t$ seconds), it also knowns the expected assimilation profile, since the concentration of molecules around the RX linearly depends on the number transmitted ones. In addition, the authors of [29] have proved that upon transmitting a burst of $Q$ molecules, the number of received molecules for large times is well approximated by

$$A_{RX}(R_{RX}, d, Q) = \frac{r_{RX}}{d} \frac{R_{RX} r_{r,rx}}{R_{RX} r_{r,rx} + \pi r_{RX}}. \quad (1)$$

Clearly, any other transmission law in different sistuations could be used. The only requirement is that the receiver is aware of it. Since molecular communications are characterized by high latency and high response times, we keep the linear increase, as in [11], to cope with slow reaction rates. The idea is to estimate the assimilation curve at run-time, and to detect congestion when a significant decrease with respect to the expected value is observed. Such a decrease has to be large enough to avoid taking natural fluctuations as congestion due to noise sources [30], [36], and enough small to be early in detecting the phenomenon. In addition, since in general the estimation of a rate value is often noisy, we prefer to take our decisions on the overall number of assimilations (i.e. the overall stimulus received, which is the integral of the rate profile), which is expected to be quadratic. Thus, the number of assimilations can be expressed as

$$N(t) = N_0 + a(t - t_0)^2, \quad (2)$$

where $N_0$ and $t_0$ are the initial offset values of type $S$ molecules and relevant time. In our algorithm, $N_0 = \zeta_{RTT}$ and $t_0 = t_{\zeta_{RTT}}$. What is missing is the estimate of the parameter $a$ of the assimilation curve, which may depend on different factors, such as trafficking time, number of receptors, distance between TX and RX, unexpected attenuation phenomena, and noise sources. We make use of a least squares estimate of $a$, that in the considered case turns out to be quite simple:

$$a = \frac{\sum_{j=1}^{N_s} (N(t) - N_0)\left(t - t_{\zeta_{RTT}}\right)^2}{\sum_{j=1}^{N_s} \left(t - t_{\zeta_{RTT}}\right)^4}, \quad (3)$$

where the number of the samples is equal to $N_s = \lfloor pRTT/T_w \rfloor$, and $p$ is the fraction of the initial RTT duration used to carry out the estimation. We selected a variable duration of the estimation time, since it is more adaptive to the working context. In particular, binding it to the RTT value allows taking into account the distance between TX and RX. We have verified numerically that, up to RTT seconds since the RTT estimation, congestion has still not occurred. The overall machinery, including all operations carried out by the RX, is illustrated in Fig. 1. As shown in the figure, the RX has to make use of a number of samples of $a$ to implement (3), and only when $a$ is estimated it starts detecting possible congestion conditions.

As for this last point, we make use of two different functions, $N_{exp}(t)$ and $N_{stop}(t)$, to determine congestion and stop conditions. The first is equal to

$$N_{exp}(t) = N_0 + \beta a (t - t_0)^2, \quad (4)$$

where $\beta$ is the tolerance coefficient used to discriminate between congestion condition and random fluctuation of absorbed molecules, whereas the offset values depend on the selected throttling rate algorithm. They are function of the values $N_c$, $t_{prev}$, and $t_c$ estimated as shown in Fig. 1, and will be illustrated in the next section. As for the stop condition, the relevant function is equal to

$$N_{stop}(t) = N(t) + 2a(t - t_0)((\|STOP\| - 1)T_S + RTT), \quad (5)$$

which represents a linearization of the estimated curve (2) at time $t$, in order to take into account possible overload conditions occurring during the transmission of the STOP signal. The quantity $\|\cdot\|$ represents the number of symbols of a given sequence. The interval $t_{lock} - t$ represents the time needed to transmit the CONTROL message. During this interval all control operations at RX are frozen, since it cannot transmit anything until the control sequence has ended.

We remark that the law used here has the objective of an *early* congestion detection, differently from the one sketched in [11], which is less robust and designed to detect only heavy overload conditions. A further observation is that, even choosing an optimal rate in advance and keeping it constant for all the transmission time is tricky, since the actual distance between nanomachines is difficult to estimate accurately. Also the number of receptors is hardly known in advance, especially in bio-hybrid nodes [2], since it may be increased or decreased at runtime depending on the internal processing of the received stimulus; these processes are known as upregulation and downregulation, respectively [37]. Finally, the presence of either inhibitory molecules in the surrounding environment [38] or any other absorbing nodes [39] would alter the expected number of assimilated particles as well. Instead, the proposed adaptive approach learns the assimilation

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2707597, IEEE Access



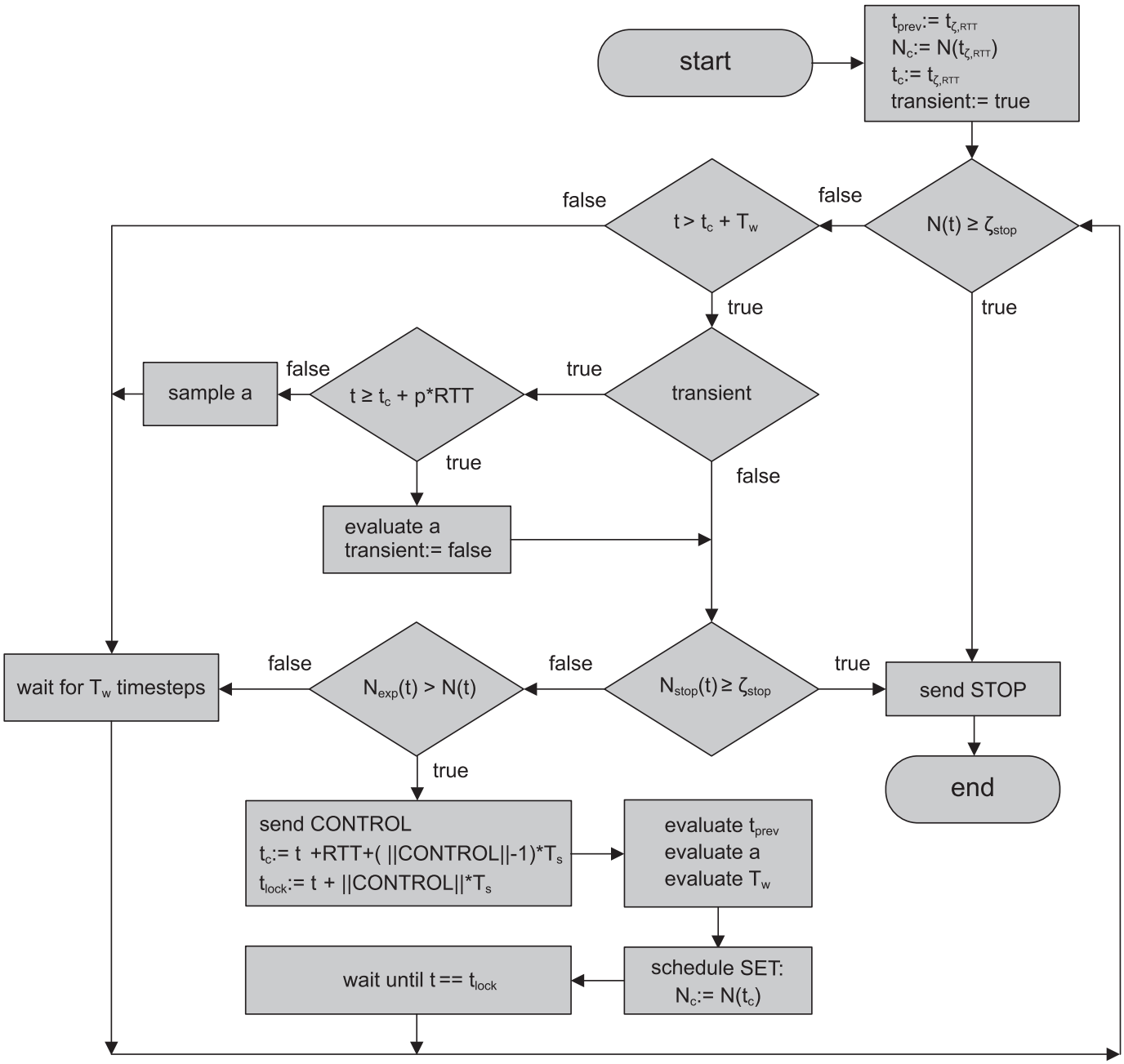


Fig. 1: Flow diagram of the operations executed by the RX.

curve blindly with respect to all these parameters, thus largely simplifying the system design.

## V. Congestion control scheme

In this section, we review four rate control algorithms, which can be triggered by the reception of a negative feedback (CONTROL message) sent by the RX. For each of them, we illustrate the new rate value upon feedback decoding, and also all the operation carried out at the RX in order to re-align with the new, expected assimilation profile.

### *A. Algorithm "Restart"*

This algorithm is inspired by TCP Tahoe. However, since the slow-start method of the TCP would be to unfeasible in this scenario due to the lack of positive feedbacks, we start the procedure with a scheme recalling congestion avoidance. Specifically, when the CONTROL message is received, the transmission burst is reset to 1 molecule, that is $B_{TX} = B_{0,TX}$. At the RX side, this implies that the parabolic estimation restarts from the minimum of the parabola. In more detail, the effect of the control message will be visible at the RX at time $t_c = t + (\|CONTROL\| - 1)\,T_S + RTT$ defined in Fig. 1 and thus $t_0 = t_c$. The new offset is estimated as the number of absorbed molecules at that time, that is $N_0 = N_c = N(t_c)$. The value of $t_c$ takes into account the duration of transmission of control message, and the RTT (see the definition in section III). Note that $t_{prev}$ in Fig. 1 is not used here. The rationale of this algorithm is to allow decongesting the RX, by restarting

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2707597, IEEE Access



transmission from the beginning.

### B. Algorithm "Restart - Double"

This variant, in addition to performing the same operation of the algorithm "Restart", doubles the transmission time $\Delta t \leftarrow 2 \times \Delta t$. As a consequence, we also double the observation time $T_w \leftarrow 2 \times T_w$ at the receiver, which has to be larger than the transmission time in order to correctly estimate the parabolic profile. The values of $N_0$ and $t_0$ are updated as in the previous case, but doubling the transmission time has an additional effect on the parameter $a$, which has to be updated accordingly. In particular, if in a time interval $T$ the number of transmission bursts are $n_T = T/\Delta t$, starting from $B_{0,TX}$, then the number of transmitted molecules is simply $N_T = n_T (n_T + 1)/2$. When the value of $\Delta t$ is doubled the number of transmitted molecules in the same interval are approximately decreased by a factor of 4, since

$$N_T = \sum_{i=1}^{n_T/2} i = \frac{n_T}{2} \frac{(n_T/2+1)}{2} = \frac{1}{4} \frac{n_T (n_T+2)}{2} \approx \frac{N_T}{4}. \quad (6)$$

Clearly, this translates into updating $a \leftarrow a/4$. The rationale of this algorithm is to allow a complete decongestion of the RX, not only by restarting the transmission process, but also by slowing the increase of the transmission rate.

### C. Algorithm "Halve"

This algorithm is presented in [11], and inspires to TCP Reno. In this algorithm, the transmission window at the TX is halved, and not re-initialized to a value of 1 molecule. Let us consider the time $t_c$ at which the effect of this action is visible at RX. Since the transmission period has not changed, the values of $a$ and $T_w$ will remain unchanged. As for the time shift $t_0$, since the rate is perceived as halved at $t_c$ and the previous starting point was estimated at $t_0 = t_{prev} = t_{\zeta_{RTT}}$, then it is shifted to

$$t_0 = \frac{t_c + t_{prev}}{2}, \quad (7)$$

and then the new value of $t_{prev}$ is set to $t_{prev} = t_0$. As for $N_{exp}(t)$, the value of $N_c$ summarizes not only the number of assimilations before the control action, but also the initial $(t_c - t_0)$ seconds of the new parabola, as mentioned above. This means that

$$N_{exp}(t) = N_c - (t_c - t_0)^2 + \beta a (t - t_0)^2, \quad (8)$$

and thus $N_0 = N_c - (t_c - t_0)^2$.

### D. Algorithm "Halve - Double"

This latter scheme is a variant of the "Halve" one, in which, similarly to "Restart - Double", upon receiving the CONTROL message the TX node doubles the transmission period $\Delta t$. This implies that $\Delta t \leftarrow 2 \times \Delta t$, $T_w \leftarrow 2 \times T_w$, and $a \leftarrow a/4$. As for $t_0$, the same comments relevant to the "Halve" approach hold and, in addition, we have to consider that the transmission rate increase is halved. Thus, according to the new transmission profile, in the time interval $(t_c - t_0)$, the TX transmits exactly *half* the number of bursts with respect to the previous configuration, and thus it simply results that $t_0$ remain unchanged ($t_0 = t_{\zeta_{RTT}}$). As for $N_{exp}(t)$, it has the same value of (8), but with the updated $a$ value.

TABLE I: Simulation parameters

| Symbol | Description | Value |
|---|---|---|
| $dt$ | Simulation time step | 20 $\mu s$ |
| $T$ | Temperature | 310 K |
| $e$ | Coefficient of restitution (partially inelastic collisions) | 0.9 |
| $\eta$ | Viscosity | 0.0011 $Kg \times (ms)^{-1}$ |
| $\beta$ | Tolerance factor | 0.95 |
| $T_S$ | Symbol time | 10 s [11] |
| $r_{RX}$ | Radius node RX | 2.5 $\mu m$ |
| $r_{TX}$ | Radius node TX | 2.5 $\mu m$ |
| $R_{RX}$ | Amount of surface receptors (node RX) | 10000 |
| $R_{TX}$ | Amount of surface receptors (node TX) | 10000 |
| $r_{c,rx}$ | Radius emitted molecules (type $R$) | 3.5 nm |
| $r_{c,tx}$ | Radius emitted molecules (type $S$) | 1.75 nm |
| $r_{r,rx}$ | Receptor radius (RX), type $S$ | 4 nm |
| $r_{r,tx}$ | Receptor radius (TX), type $R$ | 8 nm |
| $T_{traff}$ | Trafficking time | 4 s |
| $\zeta_S$ | Assimilation threshold (TX) | 100 $R$ molecules |
| $\Delta t$ | Emission time (TX) | 60 ms |
| $\zeta_{stop}$ | Assimilation threshold for STOP signal (RX) | 100000 molecules |
| $START$ | Signal pattern: START | 110 |
| $CONTROL$ | Signal pattern: CONTROL | 10 |
| $STOP$ | Signal pattern: STOP | 111 |
| $d$ | Simulated distance | 26.5 $\mu m$ |
| $B_{RX}$ | Burst for control signals (RX) | 2000 $R$ molecules |
| $B_{0,TX}$ | Initial burst (TX) | 1 $S$ molecule |
| $\zeta_{RTT}$ | Threshold for RTT estimation (RX) | 5 $S$ molecules |
| $T_w$ | Observation window default value (RX) | 200 ms |
| $p$ | Fraction of RTT used for $a$ estimation | 0.5 |

## VI. Performance evaluation

The performance evaluation of the system has been carried out by using the BiNS2 simulator. The main simulation parameters, together with their descriptions and values, are reported in Table I.

We first performed a preliminary analysis in order to select the values of $p$ parameter that allows obtaining a reliable estimation of $a$. The results are shown in Fig. 2. We found that, for very low values of the parameter $p$, the estimation of $a$ is quite unreliable, due to the low number of samples. As the value of $p$ approaches 0.5, the estimation gets stable, and the variation with $p$ becomes less significant. Thus, in order to get a reliable estimation, values in the order of 0.5-1 are recommended. As for the impact of the trafficking time on the estimation of $a$, we can see that, in average, the lower $T_{traff}$, the slightly larger the value of $a$. This can be easily explained, since, in uncongested conditions, the trafficking time does not have a strong effect. The slightly larger values for smaller $T_{traff}$ is simply due to the fact that, in a given observation time, the number of *complete* assimilations is slightly larger when the assimilation time $T_{traff}$ is smaller. The number of samples can be easily computed by considering that the observation time is 200 ms, and that the RTT value is reported on Fig. 2.b for different distances. This allows also understanding how the system dynamics evolve in different settings. In addition, we evaluated also the impact of different $T_{traff}$ and $d$ values on the estimation of $a$. We found that binding the estimation of $a$ to the value of RTT by means of $p$ is reasonable since, when the distance increases, the $S$ molecules concentration decreases, thus larger observation times due to larger RTT values do not lead to congestion. Note that the value selected for the trafficking times is realistic, since [32] reports also larger ones, in the order of many tens of seconds. In this paper, we used $p$=0.5.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2707597, IEEE Access



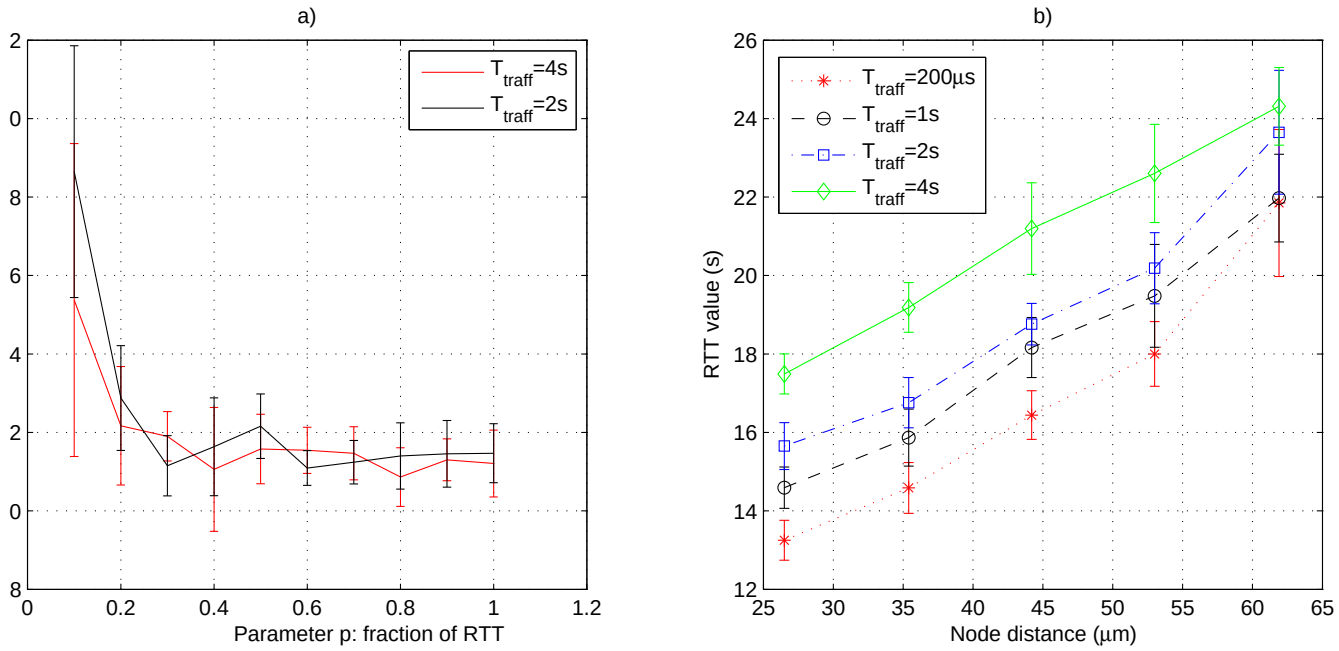


Fig. 2: a) Estimation of the value of $a$ as a function of the$p$ value, for different trafficking times. b) Dependance of the RTT time from the distance between TX and RX. 95% intervals are shown.

Now, we consider the system performance in terms of delivery efficiency and delivery time. We show some results for $T_{traff} = 4s$ and $p = 0.5$. Fig. 3 shows the assimilation and rejection profiles for the four algorithms illustrated in section V. The caption of Fig. 3 reports the meaning of the markers used in the figure.

The first comment is that the Halve rate control scheme is the one that needs the transmission of the highest number of control messages. In the shown simulation, 4 control messages are sent. The net effect is that its assimilation profile, beyond an initial part which is clearly parabolic, evolves such as a straight line, with a slope which is approximately the derivative of the parabola at the first $t_c$, which is close to the first green diamond. In addition, this "continuous" transmission of control messages, in the specific simulation shown in the figure, causes some inefficiencies in meeting the target of $\zeta_{stop}$, due to the lock effect of these transmissions on other possible actions (see also Fig. 1).

As for the Halve Double, in this case the the number of control messages is halved (i.e. 2), due to the fact that, doubling the burst transmission interval $\Delta t$, the RX node is able to decongest and needs a larger time to become congested again. In this case, the portions of three different parabolas can be well identified. As for the Restart and Restart Double scheme, both of them are characterized by two aspects: the number of control messages is just one, since they are able to completely decongest the RX node, thanks to a complete re-initialization of the transmission burst $B_{TX}$, and the full shape of the parabola is well recognizable.

Finally, a common point to non-Double schemes is that, thanks to a constant transmission interval, they are able to complete the transmission in a shorter time.

Fig. 4 shows, for the four proposed algorithms, the values of the assimilation ($\lambda_a$) and rejection ($\lambda_r$) rates, as a function of the simulation time. For all the schemes, it can be easily identified the time instant when $\lambda_a$ begins deviating from the linear increase, and thus congestion is detected.In any case, the values of the first peak for $\lambda_a$ are comparable, which are close to 1000 molecules/s, although some deviations exist due to the random nature of the assimilation process. In fact, the initial trigger of control messages is not uniform over time, since it depends on the initial, noisy estimation of $a$, and on the specific realization of the random arrival process. As for the Halve scheme (Fig. 4.a), it is evident a behavior which recalls the well known pattern of the TCP Reno. It is the control scheme that needs the transmission of the highest number of control messages. In the simulation 4 control messages are sent. The net effect is that its assimilation rate, after an initial part which is clearly linear, evolves with a decreasing slope at the first $t_c$. In addition, this "continuous" transmission of control messagescauses some inefficiencies in meeting the target of $\zeta_{stop}$, due to the lock effect of these transmissions on other possible actions (see also Fig. 1).

Instead, in the Restart approach (Fig. 4.c), it is evident a behavior similar to the TCP Tahoe, without the initial slow start effect. It is characterized by two aspects: the number of control messages is just one, since it is able to completely decongest the RX node, thanks to a complete re-initialization of the transmission burst $B_{TX}$, and the full shape of the linear absorption rate is well recognizable.

For what concerns the Double schemes, due to the increased (doubled) transmission intervals $\Delta t$, when the congestion is detected the absolute values of the assimilation rate always reaches smaller peaks. In fact, the number of transmission bursts occurring during the time necessary to transmit, propagate, and decode the CONTROL message decreases. Thus,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2707597, IEEE Access



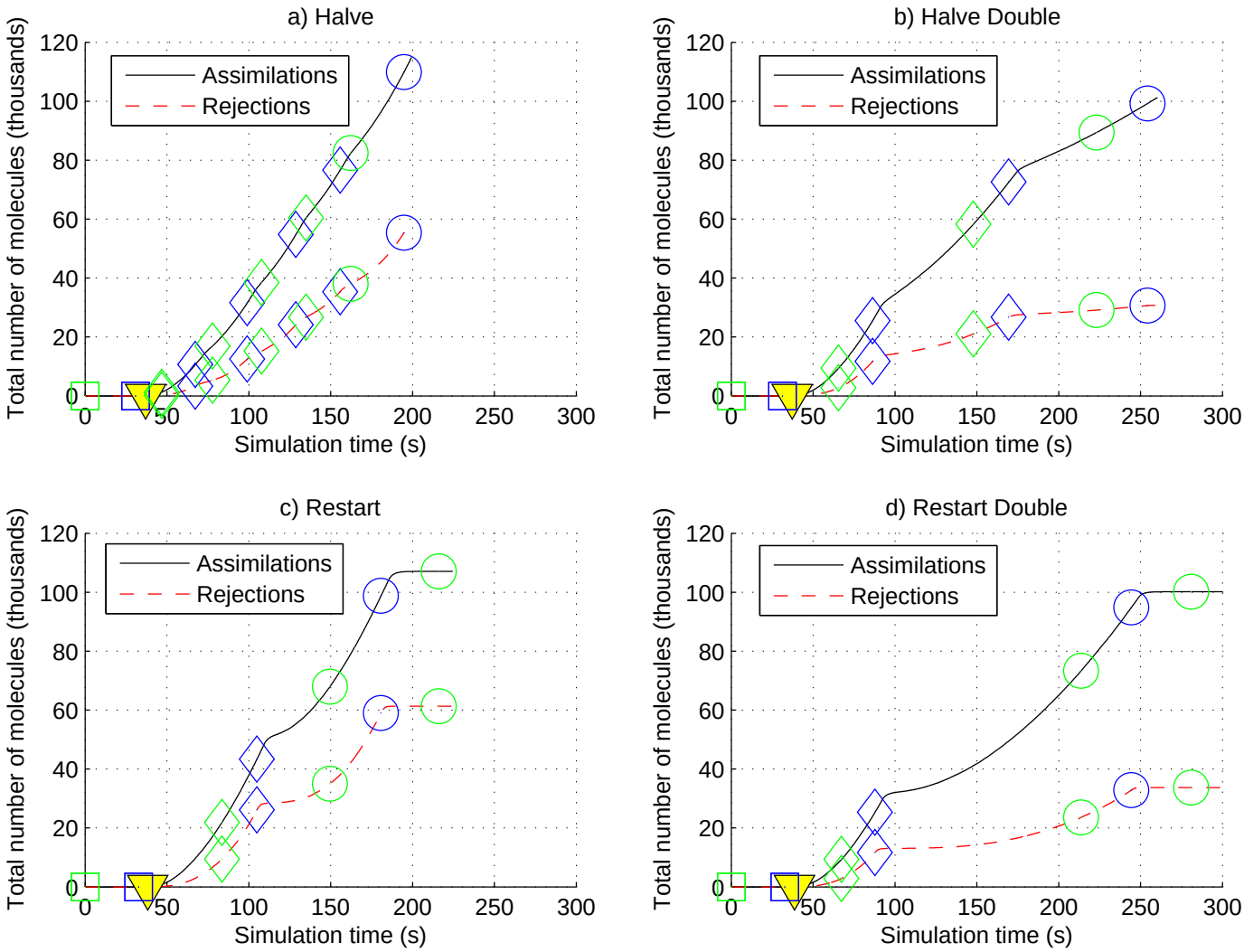


Fig. 3: Total assimilations and rejections as a function of the time. Algorithms: a) Halve, b) Halve Double, c) Restart, and d) Restart Double. Square markers represent the START signal, triangles (filled in yellow) represent the RTT estimation instant at RX, diamonds represent the CONTROL signal, and circles the STOP signal. Green markers represent the transmission instants at RX, and blue ones the relevant reception times at TX.

congestion is not exacerbated. This effect has a positive impact on the delivery efficiency. In addition, for the Restart Double scheme the same comments done for the Restart one hold.

Finally, a common point to Halve and Restart schemes is that, due to a constant transmission interval $\Delta t$, they are able to complete the transmission in a shorter time.

Fig. 5 shows, for the four proposed algorithms, the values of the rate of $\lambda_r/\lambda_a$, as a function of the simulation time. We have plotted both the instantaneous rate, behavior of which is similar to that shown in Fig. 4, and the cumulative ratio. It is interesting to note that, in the Halve scheme, the cumulative ratio is difficult to keep under control, since the transmission burst does not restart from the minimum value, and the control countermeasure allows just limiting its increase. The Restart algorithm can maintain it nearly constant, whereas the Double ones, due to the ability to reduce the transmission rate by enlarging the transmission time, succeed in lowering it in the long term, meaning that they are able to completely decongest the RX node.

Fig. 6 shows a number of performance metrics. Fig 6.a shows the delivery efficiency, defined as $\zeta_{stop}/A_{RX}(R_{RX}, d, N_T)$, where $N_T$ is the total number of emitted molecules by the TX node, as in (6), and $A_{RX}$, defined in (1), is the maximum number of molecules that the RX node can assimilate in absence of congestion upon a transmission of $N_T$ molecules and with absorbing receptors (i.e. it is an upper bound). First, we observe that the efficiency values are always below 50%. In particular, it ranges from 0.25 for the Halve scheme, to about 0.35 for Halve Double and Restart Double.

We now compare the congestion control algorithms. The Halve approach is the worst performing one in terms of efficiency, since it is the most aggressive, and forces the system to be continuously on the edge of congestion. On the other hand, Double-based approaches are the most efficient, since, by relaxing the transmission rate, they allows completely decongesting the RX node, thus improving its capability of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2707597, IEEE Access



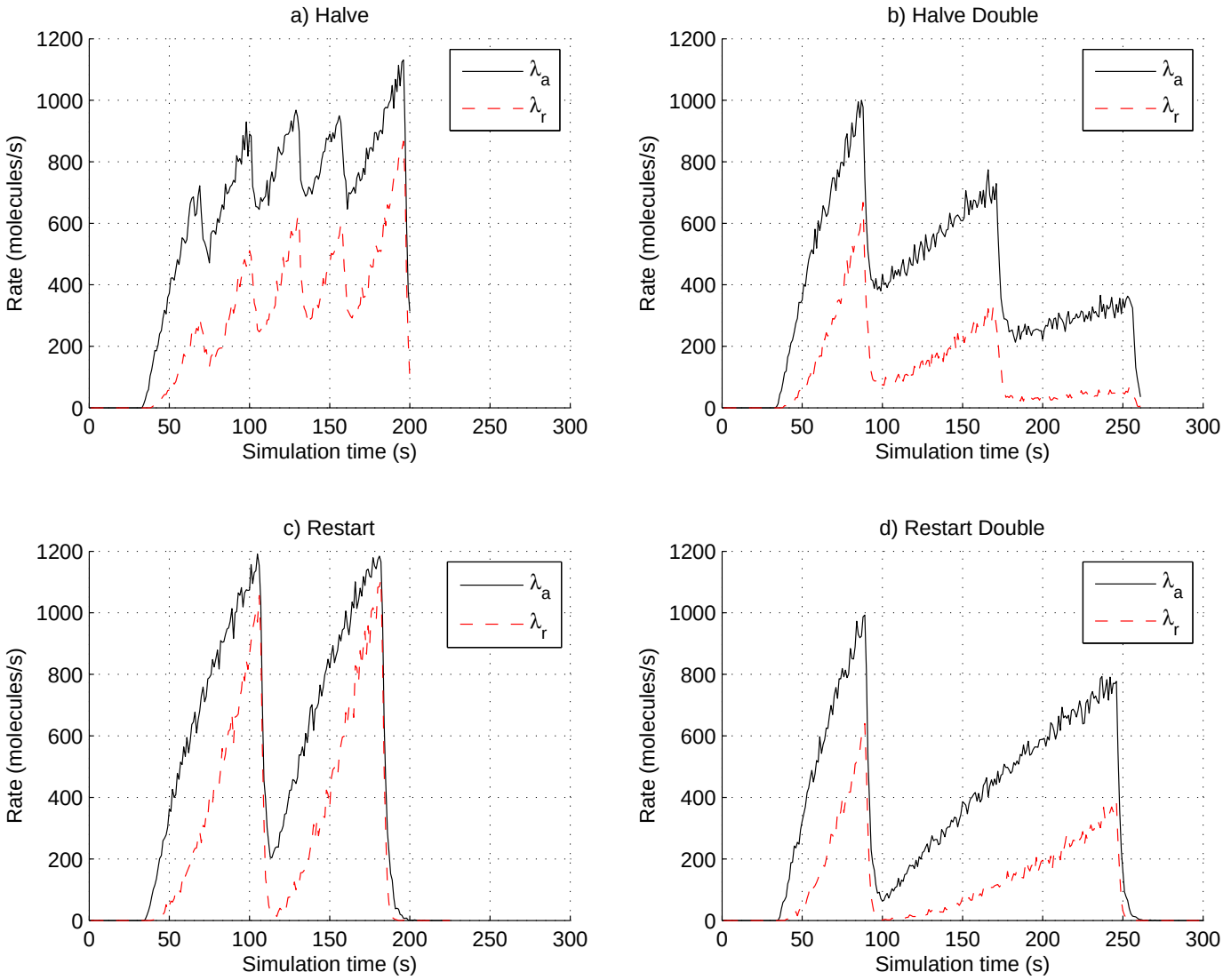


Fig. 4: Assimilation ($\lambda_a$) and rejection ($\lambda_r$) rates as a function of time. Algorithms: a) Halve, b) Halve Double, c) Restat, and d) Restart Double.

absorbing the surrounding molecules. The number of emitted carriers, shown in Fig. 6.b, is the dual metric of the efficiency. Thus, the worst performing control solution for the number of emitted molecules is the Halve, which wastes a lot of molecules since it maintains the RX node always close to saturation. Instead, Double-based approaches have similar behavior, with the best performance. Fig. 6.c shows the throughput at the receiver, and Fig. 6.d its dual counterpart, i.e. the delivery time. It appears that the Restart approach is the best performing one, having the highest throughput and, consequently, the lowest delivery time. The values achieved by the Halve approach are very close to those of the Restart one, which, however, performs better in the previous two metrics, namely efficiency and total emitted molecules. This is because, by restarting from scratch the emission rate at each congestion event, it is more efficient in decongesting the receiver with respect to the Halve approach. Obviously, the Double-based approaches are the worst performing ones with respect to throughput and delivery time. In particular, the Restart Double approach, which is slightly less aggressive than the Halve Double approach, has the lowest throughput and, consequently, the highest delivery time.

Summarizing, if we assume that all the performance metrics are equally important, we can conclude that the best choice is the Restart approach, which exhibits good values in all metrics, by balancing efficiency and throughput. However, given the simulation results, it is worth noting that the Restart approach has been tested in a slightly favorable condition, since it reaches the target $\zeta_{stop}$ with the peak rate (see Fig. 4.c). Thus, in more general conditions, it could have slightly less good performance. The less convenient approach appears to be the Restart Double, which has performance similar to Halve Double in terms of high efficiency, but worst performance in terms of throughput and delivery time.

## VII. Conclusion

In this paper, we have analyzed and compared methods for handling congestion in diffusion-based molecular communica-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2707597, IEEE Access



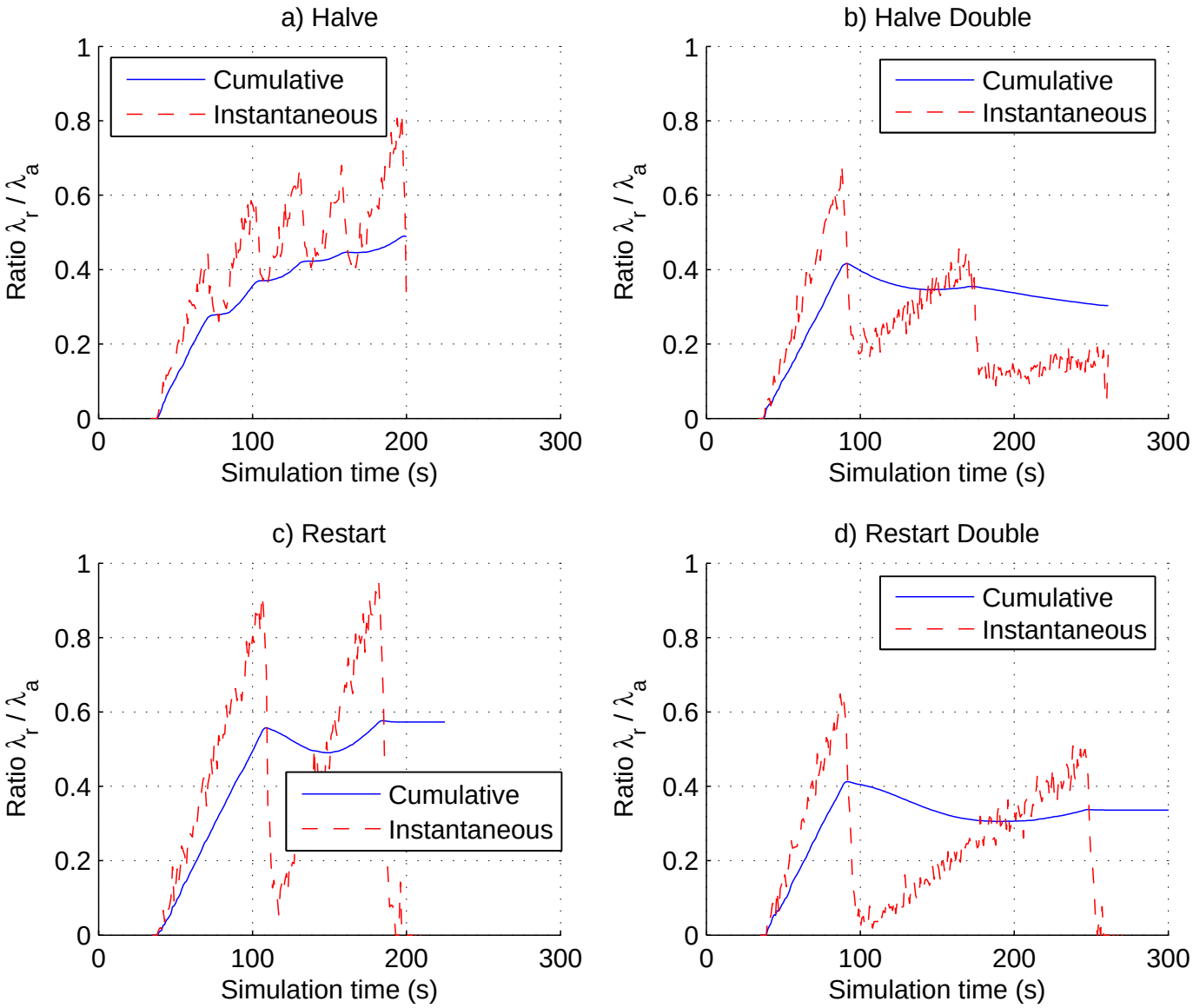


Fig. 5: Ratio of rejection and assimilation rates ($\lambda_r/\lambda_a$) as a function of time. Algorithms: a) Halve, b) Halve Double, c) Restat, and d) Restart Double.

tions. We have first proposed a simple yet effective algorithm to detect congestion in its early stages. This algorithm can be easily implemented in a device with limited memory and computing resources, as it happens in nanomachines. In addition, it needs minimal configuration, since it estimates at run-time the system behavior and adapts accordingly. Then, we have proposed and compared four different control actions, triggered by the negative feedback sent by the receiver nanomachine and implemented in the transmitter nanomachine. From our results, obtained by considering typical trafficking times, it emerges that the best control scheme is the so-called Restart one, which combines the ability to decongesting the receiver with the capacity to quickly recover towards sustained bit rates. An interesting option could be also an hybrid approach, able to relax the transmission rate by means of the so-called "Halve - Double" scheme until the need of control actions is not frequent, and then using again the "Halve" control algorithm to maintain a high throughput. Clearly, in order to implement such an hybrid scheme, it is necessary to reconsider the message encoding adopted in this analysis and previously proposed in [11]. For example, it is possible to proceed by augmenting the number of symbols and thus the round trip latency of control actions, or resorting to molecule shift keying modulation (MoSK [40]) and using two types of molecules when switching from "Halve - Double" to "Halve".

Our ongoing and future work includes the extension of the congestion control strategy in a multi-access environment, which includes multiple TX and RX nodes, the commands of which may interfere each other. Another issue, worth of investigation, is the introduction of moving nanomachines. In this case, the estimation of the assimilation profile should be repeated upon movement detection, which could be implemented by periodically repeating the estimation of the RTT with off-band signaling, which translates into the usage of an

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2707597, IEEE Access



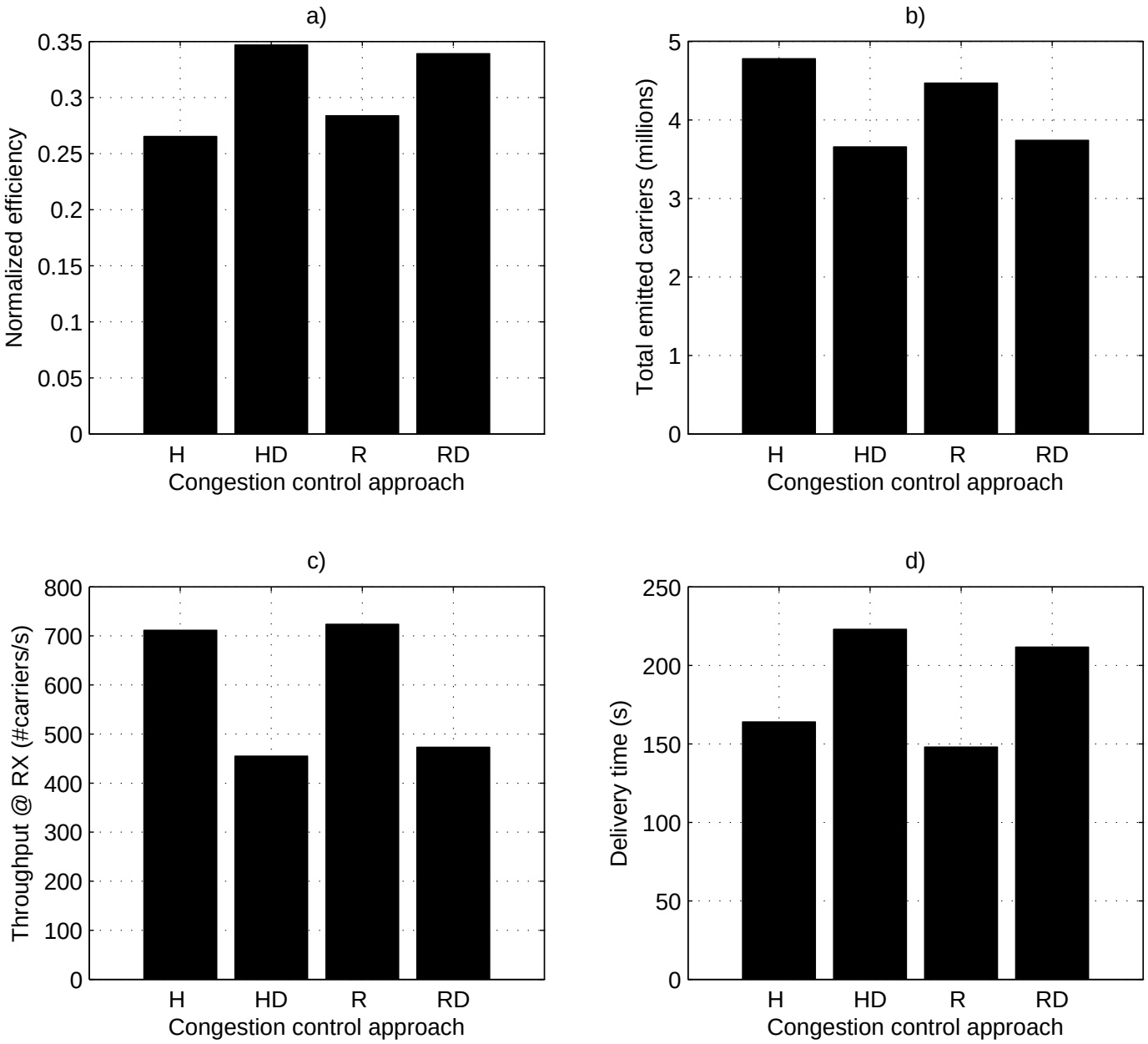


Fig. 6: Global performance of the four congestion control schemes as a function of $T_{traff}$ and $p$: a) delivery efficiency, b) total number of emitted carriers, c) receiver throughput, and d) delivery time. H means Halve, HD means Halve Double, R means Restart, and RD means Restart Double.

additional, different types of control molecules.

## VIII. Acknowledgments

This work is supported by the EU project H2020 FET Open CIRCLE (Coordinating European Research on Molecular Communications), project No. 665564, and by the MolML project funded by University of Perugia.

## References

[1] S. K. Khaitan and J. D. McCalley, "Design techniques and applications of cyberphysical systems: A survey," *IEEE Systems Journal*, vol. 9, no. 2, pp. 350–365, June 2015.

[2] I. F. Akyildiz, F. Brunetti, and C. Blázquez, "Nanonetworks: A new communication paradigm," *Comput. Netw.*, vol. 52, no. 12, pp. 2260–2279, Aug. 2008.

[3] T. Nakano, M. Moore, F. Wei, A. Vasilakos, and J. Shuai, "Molecular communication and networking: Opportunities and challenges," *IEEE Transactions on NanoBioscience*, vol. 11, no. 2, pp. 135–148, 2012.

[4] V. Loscri, C. Marchal, N. Mitton, G. Fortino, and A. Vasilakos, "Security and privacy in molecular communication and networking: Opportunities and challenges," *NanoBioscience, IEEE Transactions on*, vol. 13, no. 3, pp. 198–207, Sept 2014.

[5] L. Felicetti, M. Femminella, G. Reali, and P. Lio, "Applications of molecular communications to medicine: A survey," *Nano Communication Networks*, vol. 7, pp. 27 – 45, 2016.

[6] I. F. Akyildiz, M. Pierobon, S. Balasubramaniam, and Y. Koucheryavy, "The internet of bio-nano things," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 32–40, March 2015.

[7] U. Chude-Okonkwo et al, "Molecular communication model for targeted drug delivery in multiple disease sites with diversely expressed enzymes," *IEEE Trans. on NanoBioscience*, vol. 15, no. 3, pp. 230–245, April 2016.

[8] M. Femminella, G. Reali, and A. V. Vasilakos, "A molecular communications model for drug delivery," *IEEE Transactions on NanoBioscience*, vol. 14, no. 8, pp. 935–945, Dec 2015.

[9] L. Felicetti, M. Femminella, and G. Reali, "A simple and scalable receiver model in molecular communication systems," in *Proceedings of the 3rd ACM International Conference on Nanoscale Computing and Communication*, ser. NANOCOM'16. New York, NY, USA: ACM, 2016, pp. 39:1–39:2. [Online]. Available: http://doi.acm.org/10.1145/2967446.2967475

[10] T. Nakano, Y. Okaie, and A. V. Vasilakos, "Transmission rate control for molecular communication among biological nanomachines," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 12, supplement, 2013.

[11] L. Felicetti, M. Femminella, G. Reali, T. Nakano, and A. V. Vasilakos, "TCP-like molecular communications," *Selected Areas in Communications, IEEE Journal on*, Dec. 2014.

[12] L. Felicetti, M. Femminella, and G. Reali, "A simulation tool for nanoscale biological networks," *Nano Communication Networks*, vol. 3, no. 1, pp. 2–18, 2012.

[13] L. Felicetti, M. Femminella, G. Reali, P. Gresele, and M. Malvestiti, "Simulating an in vitro experiment on nanoscale communications by using bins2," *Nano Communication Networks*, vol. 4, no. 4, pp. 172 – 180, 2013.

[14] M. Moore, T. Suda, and K. Oiwa, "Molecular communication: Modeling noise effects on information rate," *IEEE Transactions on NanoBioscience*, vol. 8, no. 2, pp. 169–180, 2009.

[15] T. Nakano, Y. Okaie, and J.-Q. Liu, "Channel model and capacity analysis of molecular communication with brownian motion," *Communications Letters, IEEE*, vol. 16, no. 6, pp. 797–800, 2012.

[16] M. S. Kuran, H. B. Yilmaz, T. Tugcu, and B. Ozerman, "Energy model for communication via diffusion in nanonetworks," *Nano Communication Networks*, vol. 1, no. 2, pp. 86 – 95, 2010.

[17] S. Kadloor, R. Adve, and A. Eckford, "Molecular communication using brownian motion with drift," *IEEE Transactions on NanoBioscience*, vol. 11, no. 2, pp. 89–99, June 2012.

[18] K. V. Srinivas, A. Eckford, and R. Adve, "Molecular communication in fluid media: The additive inverse gaussian noise channel," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4678–4692, 2012.

[19] T. Nakano and J.-Q. Liu, "Design and analysis of molecular relay channels: An information theoretic approach," *IEEE Transactions on NanoBioscience*, vol. 9, no. 3, pp. 213–221, 2010.

[20] T. Nakano and J. Shuai, "Repeater design and modeling for molecular communication networks," in *IEEE INFOCOM Workshop*, 2011, pp. 501–506.

[21] N. Farsad, A. Eckford, and S. Hiyama, "A markov chain channel model for active transport molecular communication," *Signal Processing, IEEE Transactions on*, vol. 62, no. 9, pp. 2424–2436, May 2014.

[22] A. Guney, B. Atakan, and O. Akan, "Mobile ad hoc nanonetworks with collision-based molecular communication," *IEEE Transactions on Mobile Computing*, vol. 11, no. 3, pp. 353–366, March 2012.

[23] B. Atakan and O. B. Akan, "On molecular multiple-access, broadcast, and relay channels in nanonetworks," in *ICST BIONETICS 2010*, 2010.

[24] ——, "Deterministic capacity of information flow in molecular nanonetworks," *Nano Communication Networks*, vol. 1, no. 1, pp. 31 – 42, 2010.

[25] M. U. Mahfuz, D. Makrakis, and H. T. Mouftah, "On the characterization of binary concentration-encoded molecular communication in nanonetworks," *Nano Communication Networks*, vol. 1, no. 4, pp. 289 – 300, 2010.

[26] M. Pierobon and I. Akyildiz, "A physical end-to-end model for molecular communication in nanonetworks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 4, pp. 602–611, 2010.

[27] ——, "Diffusion-based noise analysis for molecular communication in nanonetworks," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2532–2547, 2011.

[28] H. Yilmaz, A. Heren, T. Tugcu, and C.-B. Chae, "Three-dimensional channel characteristics for molecular communications with an absorbing receiver," *Communications Letters, IEEE*, vol. 18, no. 6, pp. 929–932, June 2014.

[29] A. Akkaya, H. Yilmaz, C. Chae, and T. Tugcu, "Effect of receptor density and size on signal reception in molecular communication via diffusion with an absorbing receiver," *Communications Letters, IEEE*, vol. 19, no. 2, pp. 155–158, Feb 2015.

[30] M. Pierobon and I. Akyildiz, "Noise analysis in ligand-binding reception for molecular communication in nanonetworks," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4168 –4182, September 2011.

[31] L. Felicetti, M. Femminella, G. Reali, J. Daigle, M. Malvestiti, and P. Gresele, "Modeling CD40-based molecular communications in blood vessels," *IEEE Transactions on NanoBioscience*, vol. 13, no. 3.

[32] D. Lauffenburger and J. Linderman, *Receptors: Models for Binding, Trafficking, and Signalling,*. Oxford University Press, 1996.

[33] Y. Deng, A. Noel, M. Elkashlan, A. Nallanathan, and K. C. Cheung, "Modeling and simulation of molecular communication systems with a reversible adsorption receiver," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 1, no. 4, pp. 347–362, Dec 2015.

[34] Y. Lu, M. D. Higgins, A. Noel, M. S. Leeson, and Y. Chen, "The effect of two receivers on broadcast molecular communication systems," *IEEE Transactions on NanoBioscience*, 2016.

[35] J. Philibert, "One and a half century of diffusion: Fick, einstein, before and beyond," *Diffusion Fundamamentals*, vol. 4, pp. 6.1–6.19, 2006.

[36] M. Pierobon and I. Akyildiz, "Capacity of a Diffusion-Based Molecular Communication System With Channel Memory and Molecular Noise ," *IEEE Transactions on Information Theory*, vol. 59, no. 2, pp. 942–954, Feb. 2013.

[37] J. Feher, *Quantitative Human Physiology, An Introduction*. Boston: Academic Press, 2012.

[38] A. Noel, K. Cheung, and R. Schober, "Improving receiver performance of diffusive molecular communication with enzymes," *NanoBioscience, IEEE Transactions on*, vol. 13, no. 1, pp. 31–43, March 2014.

[39] Y. Okaie, T. Nakano, T. Hara, and S. Nishio, "Distributing nanomachines for minimizing mean residence time of molecular signals in bionanosensor networks," *Sensors Journal, IEEE*, vol. 14, no. 1, pp. 218–227, Jan 2014.

[40] H. ShahMohammadian, G. G. Messier, and S. Magierowski, "Optimum receiver for molecule shift keying modulation in diffusion-based molecular communication channels," *Nano Communication Networks*, vol. 3, pp. 183–195, 2012.