DOI: 10.1109/TNB.2014.2363002

Collision Bottleneck Throughput in Bacterial Conjugation-based Nanonetworks

Nabiul Islam, and Sudip Misra, *Senior Member, IEEE*
Indian Institute of Technology, Kharagpur, Pin: 721302, India
{nabiuli, smisra}@sit.iitkgp.ernet.in

***Abstract*—Bacterial conjugation-based nanonetwork has been recently proposed as a novel molecular communication paradigm, in which the bacteria act as carriers. This is the foundational work proposing the phenomenon of collision which occurs in the form of multi-conjugation of multiple carrier bacteria at the side of receiver nanodevice. We show the effect of this conjugation-based collision on the maximum achievable throughput of the network, using a simple graph-theoretic approach, namely, *Maximum Weight Bipartite Matching*. One of the several interesting results that emerges concerns the maximum achievable throughput, which is bounded by $\Theta(\frac{n}{p})$ in case of homogeneous nodes, where n and p refer to the total number of nodes, and the vertical layers in the network, respectively.**

***Index Terms*—Bacterial multi-conjugation collision, nanonetworks, molecular communication, and throughput.**

I. Introduction

In bacterial conjugation-based communication nanonetworks, bacteria are the carriers of information. Information are encoded as sequences of four nucleotide bases (A, T, G, C) of a plasmid, and plasmids are inserted into a bacterium genome [1]. The exchange of these information is made through the process of *conjugation* between two bacteria. Communication in bacterial conjugation-based nanonetworks is of multi-hop fashion due to the limited resource capabilities of nanodevices and carrier bacteria. At this juncture, the following questions motivate us to explore the effect of formation of multi-conjugation with several bacteria.

1) Does the process of conjugation initiate when multiple bacteria carrying different information come in the close proximity of one another?
2) Does the information content of bacteria from multiple different nanodevices change if unwanted conjugation occurs at the receiver?
3) Is *collision* in traditional electromagnetic-based wireless networks applicable in bacterial conjugation-based communication?
4) How does bacterial multi-conjugation affect the performance of these networks?

One of the interesting outcomes that emerges as a result of exploration for the answers to these questions is the bacterial multi-conjugation process affecting the maximum achievable throughput in these networks. It is noted that nanodevice and nodes are used interchangeably throughout this paper. The major *contributions* of this paper are summarized as follows:

1) Conceptualizing and modeling the collision phenomenon in the form of multi-conjugation of multiple bacteria in bacterial conjugation-based nanonetworks.
2) Developing an algorithm to determine the maximum achievable throughput, which is based on a simple graph-theoretic approach, namely Maximum Weight Bipartite Matching. The proposed *Max Throughput* algorithm runs in *polynomial* time.
3) Characterizing and analyzing the throughput of such networks. The effect of various spatial distribution of nanodevices on the maximum achievable throughput are extensively analyzed and evaluated.

II. Related Works

The existing body of research literature on bacteria nanonetworks is focussed mainly on the process of the communication mechanisms such as encoding messages in plasmids, offloading plasmids at receiver, and multi-hop routing through the process of conjugation [1], [2]. For example, Cobo-Rus and Akyildiz performed simulation-based studies on propagation delay, and end-to-end capacity in bacteria-based molecular communication [3]. On the other hand, various models for channel, noise, delay, and information-theoretic capacity expressions were proposed in *diffusion-based* molecular communication nanonetworks [4], [5]. However, our work is distinct from all the mentioned works in two significant ways. First, this is the first work to study and model the collision phenomenon with the help of multi-conjugation in bacterial conjugation-based nanonetworks. Second, we design a simple algorithm to obtain the maximum achievable throughput, while respecting the collision constraint.

On the other hand, there exists vast literature on maximum throughput in electromagnetic communication-based wireless networks, such as the seminal work of Gupta and Kumar [6], followed by others such as [7], [8]. All these works either concentrate on finding the tight asymptotic bounds on throughput, or provide approximate solutions. However, the results of these studies are inapplicable in our scenario, because bacterial conjugation-based communication is fundamentally different from electromagnetic communication in terms of characteristics of the underlying communication mechanism.

III. Bacterial Conjugation: the focal point

Bacterial conjugation is a cell-to-cell contact method by which a donor bacterium delivers the genetic information encapsulating in a plasmid molecule to another bacterium or multiple bacteria [9]. Extensive experimentation [10], [11] on bacterial conjugation in both *gram-positive* and *gram-negative*

TABLE I: Few plasmid incompatibility groups

Bacteria type	Incompatibility group	Plasmid notation
Gram negative	IncFI	F, CollbP9
	IncFII	R100, R1
	IncP	RP4
Gram positive	Inc18	pIP50

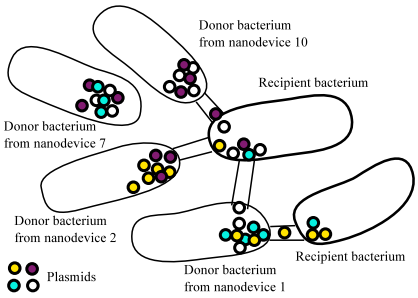


Fig. 1: Multiple conjugations.

TABLE II: Symbols and parameters used (★ indicates that the value of parameters are adopted from [17]).

Parameters and Symbols	Notations	Remarks
Bacterial density	η	$mmol/cm^3$
Density of chemo-attractant	c	2×10^{-4} $mmol/cm^3$ ★
Chemotactic coefficient	χ	3×10^{-5} cm^2/s ★
molility coefficient	β	$1.5 \times 10^{-5} cm^2/s$ ★
Consumption of attractant	λ	$1.4 \times 10^{-15} mmol/cell.s$
Average velocity of traveling band of bacteria	v	1.1×10^{-4} cm/s ★
Nodes in a vertical layer	k	Number
Total vertical layer	p	Number
Small time frame	τ	s
Maximum Throughput	Υ	KB/s
Radial distance	ζ	m
Data	φ	KB
Set of flow	Ψ	Number
Total nodes in j^{th} flow	m_j	Number
Total flows	l	Number
Angle assumed by i^{th} flow	θ_i	Number
Flow of i^{th} node	f_i	Number
Active nodes in receive layer	w_2	Number
Active nodes in transmit layer	w_1	Number

bacteria reveals that it involves the following basic sequences: constructing mating pair formation (mpf) between donor and recipient bacteria, unfolding the Origin of Transfer (*OriT*) region in plasmid of donor bacterium, and finally transferring the single strand plasmid to a recipient bacterium. In bacteria such as *Escherichia coli* (E.coli), the mpf, a multi-protein complex, renders the donor bacterium to form extracellular appendage, coined as pilus, and binds it with the recipient bacterium. After binding, another multi-protein complex, *DNA relaxosome* cuts at specific site, OriT, and passes the single strand of plasmid to the recipient bacterium.

Experimentation on multi-conjugation done by several researchers [12], [13] shows that concurrent multiple conjugations between a recipient bacterium with multiple donors is possible. The recipient bacteria can take up genetic materials from one donor followed by another. Fig. 1 illustrates the multi-conjugation. However, the coexistence of multiple plasmids inserted by multiple donor bacteria, at a recipient bacterium depends on the incompatibility group to which the plasmids belong. This is elaborated further in Section II-A.

IV. Models and Assumptions

A. Plasmids and Incompatibility group

Plasmids, which are circular double-stranded DNA molecules (distinct from *chromosomal DNA*), can be used to encoded data as sequences of its nucleotides (A, T, G and C) with the help of *watermarking* mechanism, as proposed by Gibson et al. [14]. The property of of holding multiple plasmids by an bacterium such as *E. Coli* makes an opportunity to increase network capacity, and reliability of data in such conjugation-based nanonetworks [1]. However, experimental study shows that different types of plasmids cannot coexist [15], [16], and those that do not share the communal residence are categorized as an incompatibility group. For instance, we present few groups including *IncFI* group [16] in Table I. However, the coexistence between such plasmids results in two possibilities—either one of the coresidents will be lost with high probability, known as vectoral event, or any one can be destabilized with equal probability. Therefore, it is highly likely that the process of multi-conjugation at recipient bacterium produces adverse consequence—information encoded into plasmids, which are sent through various source nanodevices, are lost. The process resembles the phenomenon of collision in traditional wireless networks.

B. Modeling of chemotactic bacteria propagation

Experimental works on bacterial motility reveals that bacteria move toward a favorable region using *run* and *tumble* mechanical processes, which is known as *chemotaxis* [10], [11], [17]. In this work, we adopt the Keller-Segal (K-S) model of chemotaxis, a well established model to depict the motility of chemotactic band of bacteria inside a capillary tube [17], [18]. The generalized version of the model [17] is given as follows:

$$\frac{\partial \eta}{\partial t} = \nabla.(\beta(c)\nabla\eta) - \nabla.(\chi(c)\eta\nabla c) + g(\eta, c) - h(\eta, c), \quad (1)$$

$$\frac{\partial c}{\partial t} = D\nabla^2 c - f(\eta, c) \quad (2)$$

where $\eta = \eta(\mathbf{x}, t)$ and $c = c(\mathbf{x}, t)$ refer to the bacterial density, and concentration of the chemo-attractant at spatial position $\mathbf{x}$ and time t, respectively. The terms $\beta(c)$ and $\chi(c)$ denote the *bacterial motility*, and *chemotactic* coefficient, respectively. The functions $g(\eta, c)$, $h(\eta, c)$, and $f(\eta, c)$ express the growth and death of bacterial cells, and the depletion of chemo-attractants, respectively. In this work, we consider $\beta(c)$=β as constant. The growth and death of bacteria cells during the propagation of band is not considered. Moreover, we assume the following characteristics: (1) The population of bacteria is propelled by chemotactic strength and bacterial diffusion; (2) Since the rate of diffusion of chemo-attractant is comparatively small, the effect of chemo-attractant's diffusion is not considered (see [18]); (3) Single chemo-attractant is considered with concentration level as $c(\mathbf{x}, t)$; and (4) Bacteria consume the chemo-attractant at a constant rate of λ.

Considering the assumptions, the revised version of Equa-

tions (1) and (2) for predicting the behavior of chemotactic band of bacteria, is written as follows:

$$\frac{\partial \eta}{\partial t} = \frac{\partial}{\partial x}\left(\beta \frac{\partial \eta}{\partial x} - \chi(c)\eta \frac{\partial c}{\partial x}\right), \tag{3}$$

$$\frac{\partial c}{\partial t} = \lambda(s)\eta \tag{4}$$

To obtain solutions of Equations (3) and (4), the travelling wave co-ordinate is taken as the new coordinate system as follows:

$$z = x - vt, \quad -\infty < z < \infty \quad \eta(x,t) = \eta(z) \quad c(x,t) = c(z)$$

where v denotes average velocity of the traveling band of bacteria. The formation of band is the effect of biased random walk due to the presence of gradient of chemo-attractant. It is noted that v determines the propagation time of the band of bacteria to reach a node. Considering the following initial conditions, and assuming the behavior of traveling band of bacteria being closed inside a capillary tube of length L [18],

$$c(x,0) = c_0(x), \quad \eta(x,0) = \eta_0(x), \quad \text{where } t > 0,\ 0 < x < L$$
$$\frac{\partial b}{\partial x} = \frac{\partial s}{\partial x} = 0, \quad \text{when } x = 0, L.$$

Equations (3) and (4) reduce to the following:

$$v\frac{d\eta}{dz} = -\beta \frac{d^2\eta}{dz^2} + \frac{d}{dz}\left(\eta\chi(s)\frac{dc}{dz}\right) \tag{5}$$

$$v\frac{dc}{dz} = -\lambda\eta \tag{6}$$

By solving differential equations (5) and (6), the closed form for density of traveling band of bacteria and concentration profile of chemo-attractant is given as follows:

$$\eta(z) = \frac{v^2 c_0}{\lambda(\chi - \beta)}\left(\frac{c}{c_0}\right)^{\chi/\beta} e^{-vz/\beta} \tag{7}$$

$$c(z) = c_0\left(1 + e^{-vz/\beta}\right)^{\frac{\beta}{\chi-\beta}} \tag{8}$$

C. Receiver-side Collision model

Concurrent transmissions from several nanodevices do not necessarily result in collision. Collision occurs only when information carrying bacteria from more than one source nanodevice arrive at a receiver nanodevice during the same timeframe. As mentioned, the attractant capability of a receiver nanodevice, which factually attracts bacteria, enables multi-conjugation and, in turn, occur collision at the receiver side. We term this as the *Receiver-side Collision* (ReSC) model.

D. Network model

Nanodevices are assumed to be distributed randomly in $\mathbb{R}^2$ space. Each nanodevice creates multiple concentric concentration gradient surfaces of different values with centers located at position of the nanodevice. We assume that each source node always has data to send, whereas each relay node aids to forward data to a sink node which is always ready to receive. An illustration of this network, as shown in Fig. 2. Moreover, we assume that a receiver node takes certain time as τ_1 for extracting a message from the conjugation process, and

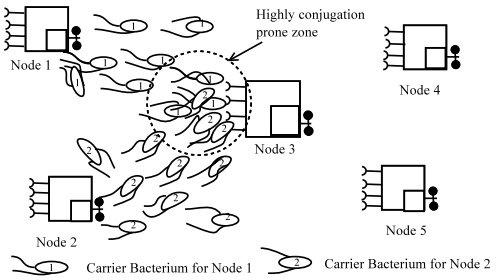


Fig. 2: Illustration of network architecture

decoding it, which is coined as the *conjugation-and-decoding* time. Once the conjugation and decoding process completes, the receiver nanodevice takes some time as τ_2 (termed as *configuration* time) for creating concentric concentration surfaces for further reception. Relay or source nodes are active for transmission for fixed amount time as τ_3. For simplicity, τ_3 is taken as equal to τ_1. We also assume that scheduling among nodes is precisely controlled by a central entity.

V. Problem Formulation

We seek to study the effect of multi-conjugation-based collision on maximum network throughput in bacterial conjugation-based nanonetworks. In other words, how much maximum data could be injected to the network such that the injected flows reach the sinks, while respecting collision constraint at the receivers – namely ReSC model, as discussed in Section III-B. At this juncture, it may be clarified that the flow or rate of flow from a source nanodevice to a receiver nanodevice is determined by the amount of data carried by the bacteria, and the time taken by the bacteria to reach the receiver nanodevice. This can be obtained from using the Equations (7) and (8), and the knowledge of velocity of the traveling band of bacteria. Therefore, in bacterial nanonetworks consisting of multiple source and sink nanodevices, the maximum achievable network throughput problem is essentially a flow optimization problem.

The network is modeled with a graph $G = (V, E)$, where V is the set of vertices representing the nodes of the network, E is set of edges representing the communication links between the nanodevices, each having capacity $C(u, v)$. If f_i^{jk} is the flow for commodity i, which is produced for the source and destination pair (j, k), then the optimization problem can be formulated as *multi-commodity flow* problem given as follows:

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} f_i^{jk}, \quad j \in S(V), k \in T(V) \\ \text{subject to} \quad & \sum_{i=1}^{n} f_i(u,v) \le C(u,v), \quad \forall i \in n \\ & f_i^{in}(u) = f_i^{out}(u), \quad u \in (V - S(V) \cup T(V) \end{aligned} \tag{9}$$

with an additional constraint that concurrent flows at any intermediate node is not permitted. The notations $S(V)$ and $T(V)$ denote the set of source and sink nodes, respectively, whereas $f_i^{in}(u)$ and $f_i^{out}(v)$ refer to the total input and output flows at any intermediate node, respectively. The quality of solution produced by *multi-commodity* is intractable [7].

Therefore, we propose a simple graph-theoretic approach, namely *Maximum Weighted Bipartite Matching* [19], which is not only tractable and but also able to provide sufficient analytical insights on the maximum throughput problem. It is noteworthy that the collision model, ReSC, is well-fitted with the graph-theoretic *matching* problem.

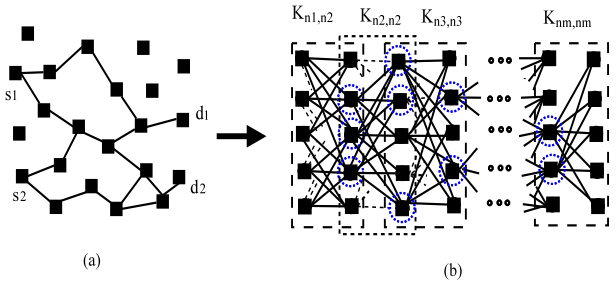


Fig. 3: Mapping to bipartite weighted graph.

A. Mapping to maximum weighted bipartite matching

The maximum throughput problem in bacterial conjugation-based nanonetwork is modeled as multiple bipartite weighted graphs connected serially, and then finding matchings of maximum weights in these graphs. The logical partitioning of the network into several bipartite graphs is shown in Fig. 3, where data are transferred from one bipartite graph to another until the data reach the sink nodes. In adopting this transformation, we do not strip the complexity of the problem, because, generally in sensor networks, sensor data produced in some locations are transmitted, via relay nodes, to sinks, which are situated at distant locations [7]. Therefore, we assume that the source nodes are located on the left side of the networks, and sink nodes on the right side, as shown in Fig. 3. The parameter *weight* represents the amount of flow to be transmitted form one node to another. Formally, the maximum throughput problem reduces to the problem of finding a *matching vector* $\mathcal{M} = [M_1, M_2, \cdots M_{n-1}]$ such that the following condition holds:

$$\begin{aligned} \underset{\mathcal{M}}{\operatorname{argmax}} \quad & \Upsilon = \sum_{Z} \left(w_1(e_1) + w_2(e_2) + \cdots + w_n(e_n)\right) \\ & Z \equiv e_1 \in M_1, e_e \in M_2, \cdots, e_n \in M_n \end{aligned} \tag{10}$$

Definition 1. ***Bipartite graph***: The set $V(G)$ can be partitioned into two independent sets X and Y such that edges follow the rule $E(G) : X \rightarrow Y$.

Definition 2. A ***matching*** is a set of non-loop edges in a graph $G = (V, E)$, in which no two edges share a common vertex.

Definition 3. ***Maximum weight Bipartite matching*** refers to the problem of finding maximum weight of a matching of a given bipartite graph $G(V, E)$ with two bipartition (X, Y), and the weight function defined as $w : E \rightarrow \mathbb{R}$. Further, a *perfect* matching render every vertex saturated by that matching.

B. Flow assignment

The term *weight* in Bipartite matching graph refers to flow of a link in bacterial nanonetworks. Let us consider a node i in the partite set X containing source nodes. From node i, the amount of flow destined to a node j belonging to other partite set Y (where the set Y consists of receiving nodes) depends on the Euclidean distance, and the amount of bacteria in the traveling band, which, in turn, is determined by using Equations (7) and (8). If the attractant capability of all nodes of set Y is the same, then Euclidean distance and velocity of traveling band of bacteria are the key determinants of the total amount of flow reachable at node j. To obtain a closed formula for the amount of data reachable at node j in a partite set Y, having Euclidean distance as $\zeta = \sqrt{x^2 + y^2}$ (where (x, y) is co-ordinates of the node j) from a node i in partite set X, we adopt the logical deduction as follows. If we consider φ amount of data (e.g., in the order of KB) being encoded in η amount of bacteria, which is to be transmitted by the source node i, the net amount of flow reachable to node j in the conjugation-and-decoding time τ_1 is governed by the following equation:

$$f = \frac{\varphi}{\tau} \times \tau_1 = \frac{\varphi}{\frac{\tau}{\tau_1}} \tag{11}$$

where $\tau = \frac{\zeta}{v}$ is the time required by η amount of bacteria, as a whole, to reach to node j, and $\frac{\tau}{\tau_1} = \iota$ refers to the *flow-reduction* coefficient.

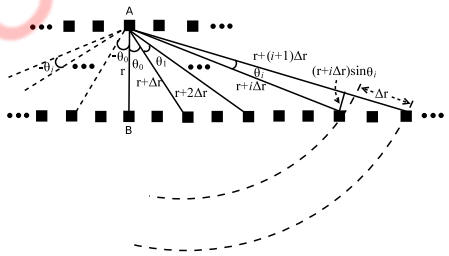


Fig. 4: Diagram of layers of sending and receiving nodes

VI. Maximum Achievable Throughput

Theorem 1. *The Complexity of the Max Throughput algorithm is bounded by* $O\left(\frac{n^3}{p^2}\right)$

Proof. The algorithm *Max Throughput*, as presented in Algorithm 1, to determine the maximum achievable throughput reveals that it is basically the *Hungarian Method* for finding a matching of maximum weight in Bipartite graph, which runs for multiple times in multi-stages. For each stage, a bipartite graph of maximum nodes $\frac{2n}{p}$ is given as input. The Hungarian Method runs in $O(n^3)$ time, where n is the number of nodes in a Bipartite graph [19]. Therefore, each stage of the Max Throughput algorithm requires time complexity as $O\left((\frac{n}{p})^3\right)$. The process of weight assignment in Line no. 3 of Algorithm 1 requires $O\left((\frac{n}{p})^2\right)$ time, as it involves each node in a partite set X to visit every node in other partite set Y. So, in every stage the time complexity is bounded by $O\left((\frac{n}{p})^3\right)$. Since, the number of stages is bounded by $(p-1)$ stages, so the time complexity is bounded by $O\left(\frac{n^3}{p^2}\right)$. □

Algorithm 1 Max Throughput

Require: 1. Set of source nodes S, sink nodes T.
2. Given the network as a $k \times P$ grid.
Ensure: Maximum throughput, Υ.

$\Upsilon \leftarrow 0$, $X_1 \leftarrow |S|$
for $i \leftarrow 2, p$ **do**
  $W_{|X_{i-1}|\times|Y_i|} \leftarrow w_{ij}$ ▷ Assigning weights to links
  $w(M) \leftarrow$ HUNGARIAN$(X_{i-1}, Y_i, W_{|X_{i-1}|\times|Y_i|})$
  $\Upsilon = \Upsilon + w(M)$
  $X_{i-1} \leftarrow Y_i$
end for
Υ
procedure HUNGARIAN$(X, Y, W_{|X|\times|Y|})$
  Generate an initial feasible labelling l in G_l
    ▷ G_l is equality subgraph [1]
  Find a matching M in E_l
    ▷ $E_l = \{(x, y) : l(x) + l(y) = w(x, y)\}$,
    ▷ $l : (X \cup Y) \rightarrow \mathbb{R}$ [1]
  if M is *perfect* matching **then**
    Stop and **return** $\sum_{e\in M} w(e)$
  else
    Choose a free vertex $u \in X$
    $R \leftarrow \{u\}$, $Q \leftarrow \emptyset$
  end if
  if $N_l(R) = Q$ **then**
    $\epsilon = \min_{x\in R, y\in Q}\{l(x) + l(y) - w(x, y)\}$
    $l_v^{new} = \begin{cases} l(v) - \epsilon & \text{if } v \in R \\ l(v) - \epsilon & \text{if } v \in Q \\ l(v) & \text{otherwise} \end{cases}$ ▷ Update labels
  end if
  if $N_l(R) \neq Q$ **then**
    Choose $y \in N_l(R) - Q$
    if y is free **then**
      Report $u - y$ as a *augmenting* path [1]
      Augment M and **go to** 12
    else if y is matched to a $w \in X$ **then**
      $R \leftarrow R \cup \{w\}$, $Q \leftarrow Q \cup \{w\}$
      go to 18
    end if
  end if
end procedure

VII. Analysis

A. Nodes in concentric concentration gradient surfaces

Let us consider a node j in a layer i, having data to inject into the next layer $(i + 1)$ (layer is synonymously referred to as partite set). Evidently, the maximum data rate achievable in the $(i + 1)^{th}$ layer, lies around its perpendicular distance (since this is the shortest among all distances measured from the node). Therefore, nodes in the $(i + 1)^{th}$ layer, which are covered by the angle θ_0, as shown in Fig. 4, get the maximum flow (data and flow are used here synonymously). We term this flow as the ψ_1 *flow*. Similarly, the nodes falling in θ_{i+1} individually get flow which is less than that of ψ_i, because of the increasing distances. Therefore, from a node j, in a vertical layer i (alignment of the position of nodes in a partite set resembles the vertical layer), the flows that can be injected into the next layer $(i + 1)$ can be partitioned into a set of l discrete flows, as given below:

$$\Psi_j = \{\psi_1, \psi_2, \psi_3, ... \ \psi_l\} \quad \text{where } \psi_1 > \psi_2 > \psi_3, \ ... \ > \psi_l \tag{12}$$

where ψ_j corresponds to the flow that each of the m_{ih}^j nodes in layer $(i+1)$ gets from a node h of the i^{th} layer. It may be noted that m_{ih}^j nodes lie either on the line segment subtending the angle θ_i or spread on two line segments, one subtending the angle θ_i, and the other subtending the angle $-\theta_i$, as shown in Fig. 4. Importantly, angles $-\theta_i$'s may not be existed, if the nodes are assumed to take the position in the end of a vertical layer. Let us consider the nodes lying in an annular space of width Δr bounded with the inner radius $(r + i\Delta r)$ and outer radius $(r+(i+1)\Delta r)$, as shown in Fig. 4 is m_j (for the purpose of generality, we use notation m_j instead of m_{ih}^j). The length of the line segment subtending the angle θ_i is given as follows:

$$s \approx \sqrt{(\Delta r)^2 + ((r + i\Delta r)\sin\theta_i)^2} \tag{13}$$

where Δs is the spatial distance between two nodes in a vertical layer. Therefore, the total number of nodes lying in the line segment, denoted as $m_{\theta i}$, is given as follows:

$$m_{\theta i} = \left\lfloor \frac{s}{\Delta s} \right\rfloor = \left\lfloor \frac{\sqrt{(\Delta r)^2 + ((r + i\Delta r)\sin\theta_i)^2}}{\Delta s} \right\rfloor \tag{14}$$

Hence, if $-\theta_i$ is present, m_j can be written in following form:

$$m_j = 2m_{\theta i} = \left\lfloor \frac{\sqrt{(\Delta r)^2 + ((r + i\Delta r)\sin\theta_i)^2}}{\Delta s} \right\rfloor \tag{15}$$

Theorem 2. *If the sufficient condition for maximum achievable throughput is to choose the flow f_i^j by an active node i in a receiving layer j such that*

$$f_i^j = max(\Psi_j), \tag{16}$$

then the maximum achievable throughput is given by

$$\Upsilon = \sum_{j=1}^{p-1}\sum_{i=1}^{m} f_i^j \tag{17}$$

where p is the total number of vertical layers, and m the total number of active nodes in the j^{th} receiving end, which get flows from the immediately previous sending layer.

Proof. Let us consider a sender node s_i in the i^{th} sending layer. The total number of distinct flows from the node can be categorized into a set of l flows. Since l flows follow the order $\psi_1 > \psi_2 > \psi_3, ... > \psi_l$, as governed by Equation (12), a node r_j at the receiving end must lie on the shortest distance line from the node s_i to get the maximum flow. As we assumed that each node in the receiving end is active (which means it is ready to receive incoming data), only those nodes lying in the region in which each point is either of shortest or of nearly shortest distance from node s_i, will be selected. Clearly, the selected flow is $f_i^j = max(\Psi_i)$. This is true for all nodes in the sending layer. Since, there exists p layers, then the maximum

[1]For maintaining brevity of the paper, we refer [19]

achievable throughput (flow) is the summation of all flows of the $(p-1)$ sending layers. □

Theorem 3. *For the homogeneous case, where each receiver has the same attractant capability, the upper bound on the maximum achievable throughput is bounded by $\Theta(\frac{n}{p})$, where p is the total number of vertical layers.*

Proof. As we know that the maximum number of nodes in a vertical layer is $\frac{n}{p}$, and each of those nodes is active in the homogeneous case, then all the flows of any layer i can be injected into the next $(i+1)^{th}$ layer. Therefore, the achievable maximum throughput is bounded by $\Theta(\frac{n}{p})$. □

B. Vulnerable layers

It is observed that the value of the aggregate throughput depends on the number of active nodes and their position in each layer. Let us suppose that there exists w number of active nodes in each layer. Among all $\binom{k}{w}$ possible groups of active nodes, there might exist some group of nodes in two consecutive layers in such way that the injected flows from one layer to the other layer reduce abruptly. We term this type of existence of spatial distribution of nodes between two partite sets as the *vulnerable* layer. The impact of vulnerable layers on maximum network throughput is discussed in the following.

C. Active nodes are uniformly distributed

Let us consider that w_1 and w_2 active nodes are uniformly distributed in the transmitting and receiving ends of a vulnerable layer, respectively. If spatial distribution of w_1 nodes teams with w_2 nodes, according to the respective perpendicular distances, each of the w_2 nodes gets maximum flow, which is termed as the *lobe-1* flow. It represents the maximum flow as ψ_1, as given in Equation (12), and the name *lobe* resembles the *front*, and *back lobe* of electromagnetic antennas. The sufficient condition of Theorem 2 satisfies when w_1 nodes individually align along lobe-1 flows, i.e., an *injective mapping* from the set w_1 to w_2, based on the respective perpendicular distances among the nodes is constructed. Let E be the event of mapping of distinct lobe-1 flows. The probability of this event can be formulated as follows:

$$\mathcal{P} = \sum_{1}^{z_1} \mathcal{P}(E|Y_j)\mathcal{P}(Y_j) \tag{18}$$

where Y_j is a random variable denoting the possible orientation of active nodes in the receiving layer, and $z_1 = \binom{k}{w_2}$ is the total number of possible ways w_2 nodes take the orientation in the receiving layer. Since, the distribution of active nodes in both the transmitting and the receiving layers is independent and each of the possible orientations is equally probable, for each j, $\mathcal{P}(E_{z_2}|Y_j)\mathcal{P}(Y_j)$ can be expressed as:

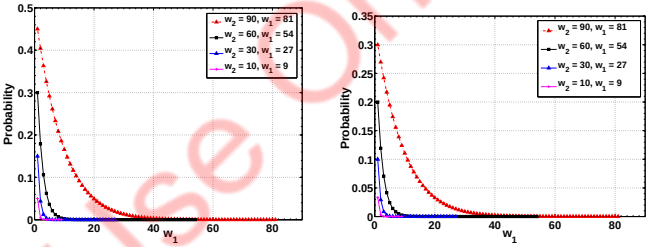
$$\mathcal{P}(E|Y_j)\mathcal{P}(Y_j) = \frac{\binom{w_2}{w_1}}{z_2 l}.\frac{1}{z_1} \tag{19}$$

where $z_2 = \binom{k}{w_1}$ is the total number of possible ways w_1 can assume orientations in the sending layer, whereas l is defined earlier as the total number of flows a node can assume. From Equations (18) and (19), we have the following form:

$$\begin{aligned}\mathcal{P} &= z_1 \cdot \frac{\binom{w_2}{w_1}}{z_2 l} \cdot \frac{1}{z_1} \\ &= \frac{1}{l} \cdot \frac{w_2.w_2(1-\frac{1}{w_2})\cdots w_2(1-\frac{w_1-1}{w_2})}{k.k(1-\frac{1}{k})\cdots k(1-\frac{w_1-1}{k})}\end{aligned} \tag{20}$$

Applying first order approximation of Taylor series expansion of e^x as, $e^x = 1+x$ when $x << 1$, Equation (20) can be written as follows:

$$\begin{aligned}\mathcal{P} &\approx \frac{1}{l} \cdot \frac{w_2^{w_1}.e^{-\frac{1}{w_2}}.e^{-\frac{1}{w_2}}\cdots e^{-\frac{w_1-1}{w_2}}}{k^{w_1}.e^{-\frac{1}{k}}.e^{-\frac{2}{k}}\cdots e^{-\frac{w_1-1}{k}}} \\ &= \frac{1}{l} \cdot \left(\frac{w_2}{k}\right)^{w_1} \cdot e^{-\frac{w_1}{2k}\cdot(w_1-1)(\frac{k}{w_2}-1)}\end{aligned} \tag{21}$$



(a) No. of distinct flows, l=2 (b) No. of distinct flows, l=3

Fig. 5: Probability of matching with distinct lobe-1 flows

Proposition 1. *For certain realization of w_2, probability $\mathcal{P}$ assumes the maximum value when w_1 assumes as follows:*

$$w_1 = \frac{2k\log(\frac{w_2}{k}) + \frac{k}{w_2} - 1}{2(\frac{k}{w_2} - 1)}$$

Proof. Taking the logarithm of Equation (21), we have:

$$\log(\mathcal{P}) = \log l + w_1 \log(\frac{w_2}{k}) - \frac{w_1}{2k}(w_1-1)(\frac{k}{w_2}-1) \tag{22}$$

After differentiating Equation (22) w.r.t. w_1, and subsequently setting the derived equation to zero, the critical value is

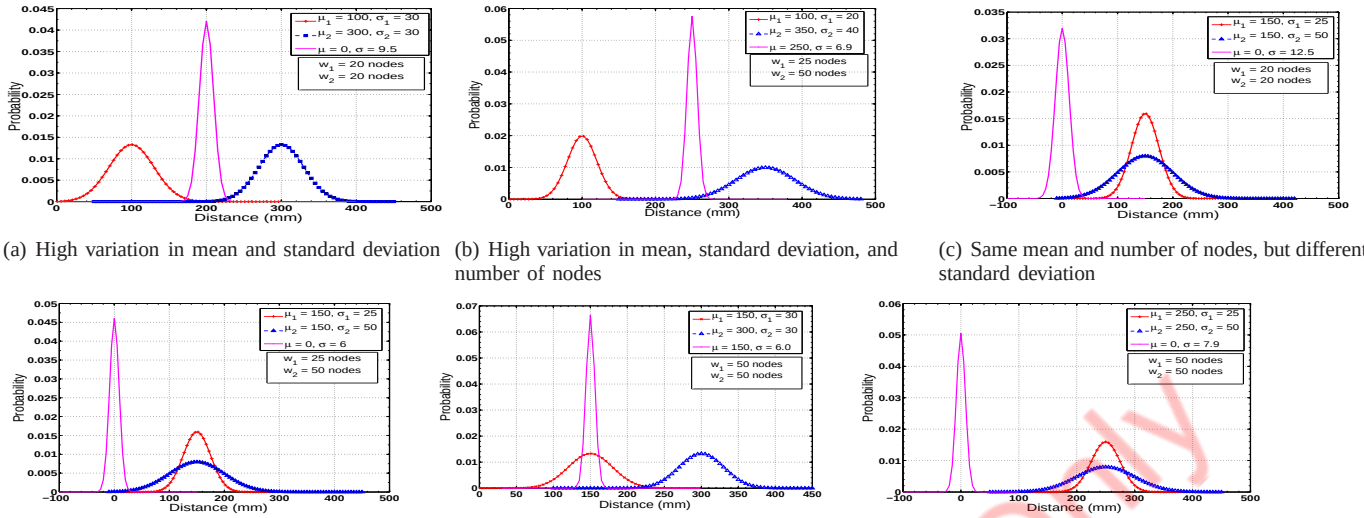
$$w_1 = \frac{2k\log(\frac{w_2}{k}) + \frac{k}{w_2} - 1}{2(\frac{k}{w_2} - 1)} \tag{23}$$

By double differentiating Equation (22) w.r.t. w_1, we get:

$$\frac{d^2}{dw_1^2}(\log\mathcal{P}) = -\frac{1}{k}(\frac{k}{w_2}-1) \qquad < 0 \tag{24}$$

Since $k > w_2$, the second derivative is less than zero. So, the maximum value is achieved at $w_1 = \frac{2k\log(\frac{w_2}{k})+\frac{k}{w_2}-1}{2(\frac{k}{w_2}-1)}$. □

Fig. 5 shows the relationship between parameters such as the distinct flow, l, the number of nodes in the sending layer, w_1, and the number of nodes in the receiving layer, w_2, and reveals that as the distinct flows supported by each node increase, the probability values decrease. It is due to the fact that the probability describes the deviation from satisfying the sufficient condition constraint, which is outlined in the first part of Theorem 1.



(a) High variation in mean and standard deviation (b) High variation in mean, standard deviation, and number of nodes (c) Same mean and number of nodes, but different standard deviation

(d) Same mean, but high difference in standard deviation and number of nodes (e) Different mean, but same standard deviation (f) Both mean, standard deviation, and number of nodes same

Fig. 6: Various combinations of means, standard deviation, and number of nodes between two sets of nodes.

D. Active nodes are Gaussian distributed

Let us consider that w_{i-1} and w_i are the number of nodes in the $(i-1)^{th}$ and i^{th} layers, and are spatially Gaussian distributed with means and variances $(\mu_{i-1}, \sigma_{i-1}^2)$ and (μ_i, σ_i^2), respectively. Gaussian distribution is chosen because it is a naturally occurring probability distribution, and all the other distributions may converge to it, according to the *central-limit* theorem [20]. We are interested in the distribution of distance of a randomly selected active node. The Euclidean distance between a node in the $(i-1)^{th}$ layer and another in the i^{th} layer is given as follows:

$$d = \sqrt{(x_{i-1} - x_i)^2 + \kappa^2} = \sqrt{\Delta x^2 + \kappa^2} \qquad (25)$$

where $(x_{i-1}, 0)$, and (x_i, κ) are the coordinates of nodes in the $(i-1)^{th}$ and i^{th} layers, respectively. The origin of $X-$axis of the reference system is chosen at one end of the $(i-1)^{th}$ layer, and κ is the perpendicular distance between two layers. Since the distribution of nodes of two layers is an independent Gaussian distribution, the distribution of Δx follows as $\Delta X \sim N\left(\mu_{i-1} - \mu_i, \sigma_{i-1}^2 + \sigma_i^2\right)$. According to the sampling theory [20], the distribution of two means follows the Gaussian distribution, as $X_{\mu_{i-1}-\mu i} \sim N\left(\mu_{i-1} - \mu_i, \frac{\sigma_{i-1}^2}{w_{i-1}} + \frac{\sigma_i^2}{w_i}\right)$.

1) Some insights

Evidently, a randomly chosen node falls in the lobe-1 flows, if d is minimum. Since Δx assumes probabilistic distribution, the probability of d being in lobe-1 flows can be calculated by considering $|\Delta x| \leq \varepsilon$, where ε is a small, positive quantity. Moreover, several interesting facts emerge from Fig. 6, such as: (1) Even if having the same mean and standard deviation for both sets of nodes, more number of w_2 nodes is likely to include more occurrence of the event of falling into lobe-1 flows, and (2) High variation in mean and standard deviation contributes key role to determine vulnerable layers, as shown in Figs 6(a), and 6(c).

VIII. Results

A. Simulation Design

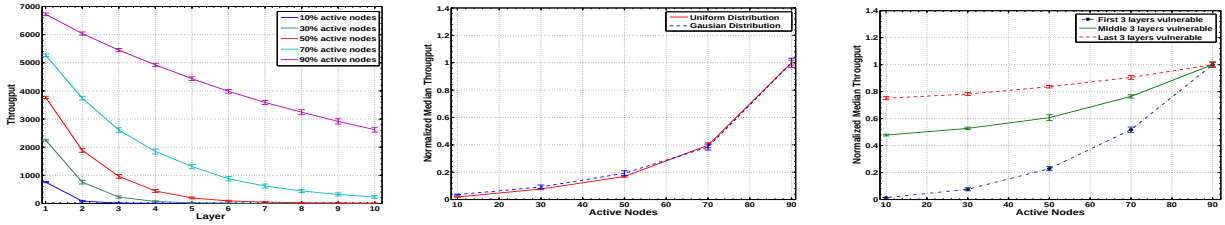
Nodes are assumed to be distributed in a grid topology with width and height taken as 10 mm. For each of the 10 vertical layers, 100 nodes are assumed. For each simulation iteration, each source node of the leftmost layer, as indicated in Fig. 3, independently generates random traffic (between 50 and 100 units), calculates the distance for each node of the second layer, and assigns flow to it according to Equation (11). Then the maximum weighted bipartite matching is executed to obtain the net injected flow reachable to the second layer. Similar process continues upto the last layer of nodes to finally determine the aggregate throughput. As simulations were repeated more than 40 times, we take the 99% confidence interval of several parameters when the cases arise.

B. Each layer contains certain percentage of active nodes

In case of uniform distribution of active nodes' positions, we observe that as active nodes increases 20% in each step from 10% upto 90% in each layer, the value of the aggregate throughput also increases, as shown in Fig 7(a). However, the injected flow from one layer to another increases substantially when an increase of 20% in active nodes occurs from 70% to 90% slab. This is attributed to the fact that active nodes are encountered in more number of lobe-1 flows. We also observed that the variation was not so noticeable in case of Gaussian distribution, as reflected in Fig. 7(b), in which the median throughputs for both the cases were compared.

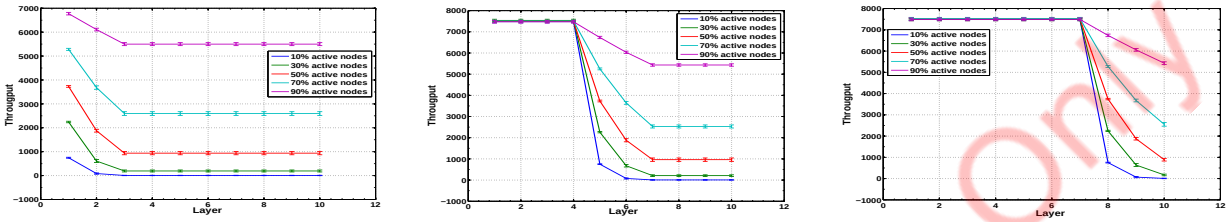
C. Few layers contain certain percentage of active nodes

Up to 30% of all the layers is considered to be vulnerable. In the simulation, the vulnerable layers are chosen in the following manner: (1) First 3 layers, (2) Middle 3 layers, and (3) Last 3 layers. The variation of active nodes is chosen similarly as described in the previous Subsection B. An important observation is that the position of a vulnerable layer does not affect the aggregate throughput reachable at the end



(a) Uniform distribution of nodes. (b) Uniform and Gaussian distribution of nodes (c) Convergence of throughputs.

Fig. 7: Throughput for different distributions of active nodes, and Convergence of throughputs.



(a) Near the layer of source nodes. (b) At the middle of network topology (c) At the end of network topology.

Fig. 8: Vulnerable layers in different locations, and corresponding throughputs.

layers, as shown in Figs. 7(c) and 8.

D. Every layer contains random active nodes

As we do not know *a priori* the exact pattern of active nodes in each layer. So, we vary the number of active nodes from 10% to 90%, randomly, in each layer multiple times. Fig. 9 shows that the throughput decreases exponentially from one layer to another, which conforms with the analytical results presented in Section VII-C.

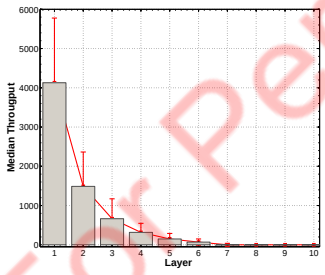


Fig. 9: Active nodes are randomly distributed.

IX. Discussion

The analysis of this work is based on grid topology. However, this restricted model does not limit the significance of our work. Rather it provides us profound insights on the achievable maximum throughput for other topologies. Since the grid topology is simple, systematic topology, the computed maximum throughput will serve as the upper bound for other more complex network topologies such as random, scale-free. Further, we considered nodes lying on different concentric concentration gradients from a receiving node. However, the impact on throughput by considering these factors such as heterogeneous concentration surface, gradient types such as *linear*, and *step*, and multipath routing is left as future works. Furthermore, it is a challenging work to determine the throughput, while honoring fairness parameter of each node.

A. BNR: The Bacterial count to Noise Ratio

Like the Signal-to-Noise Ratio (*SNR*) in traditional electromagnetic wireless networks, we argue that bacterial cell counts above a threshold limit could be used to decode a message in bacterial conjugation-based nanonetworks. This is due to two facts: (a) random motilities of bacteria, which cause some bacteria to diffuse to unintended receiver, and (b) exponential decay of bacterial cell density, as governed by Equation (7). We coin the term *Bacterial count to Noise ratio* (*BNR*) to signify SNR in such nanonetworks. Here, *Noise* includes all types of unintended cells that contribute to the bacterial count. However, SNR depends mainly on the distance between receiver and sender, whereas BNR relies on primarily the strength of *chemotactic* capability of a receiver nanodevice. In a traditional wireless network, neighbors within communication range of a transmitter receive the same messages, whereas in bacterial nanonetworks, the neighbors with high attractant capability get the maximum chunk of data.

B. Design of receiver module

We advocate the use of *biofilm-interfaced* receiver module [1], as it facilitates more bacterial conjugations. However, the cumulative replication of plasmids across a biofilm [21] poses a detrimental effect when bacteria from one of the competing source nanodevices might occupy the biofilm and capture its whole surface of the receiving module. Therefore, we envisage *microfluidics*-based microchannels [22] to be embedded in the receiving module of the nanodevice so that the development of biofilm is limited to each microchannel.

X. Conclusion

In this work, we show that the process of multi-conjugation plays the role of *collision* phenomenon in bacterial

conjugation-based nanonetworks and analyze the effect of multi-conjugation on maximum achievable throughput on such networks using a simple graph-theoretic approach, namely Maximum Weight Bipartite Matching. We propose *SNR*-like metric such as *BNR* as a prime metric to decode message at a receiver nanodevice. Thorough analysis indicates several insightful information to the quest of implementing different facets of these networks, such as the design of receiver module, efficient routing protocols, and coordination of multiple concurrent nodes. It is envisioned that nanodevices harvest energy from its environment. Therefore, how this energy-constrained environment affects the maximum achievable throughput, which is left as our future works.

ACKNOWLEDGMENT

The work of the first author was supported by Maulana Azad National Fellowship, India.

REFERENCES

[1] S. Balasubramaniam and P. Lio, "Multi-hop conjugation based bacteria nanonetworks," *IEEE T. Nanobiosci.*, vol. 12, no. 1, pp. 47–59, 2013.
[2] M. Gregori and I. F. Akyildiz, "A new nanonetwork architecture using flagellated bacteria and catalytic nanomotors," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 4, pp. 612–619, 2010.
[3] L. C. Cobo and I. F. Akyildiz, "Bacteria-based communication in nanonetworks," *Nano Commun. Netw.*, vol. 1, no. 4, pp. 244–256, 2010.
[4] M. Pierobon and I. F. 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.
[5] 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.
[6] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, 2000.
[7] C. Peraki and S. D. Servetto, "On the maximum stable throughput problem in random networks with directional antennas," in *Proceedings of the 4^{th} ACM international Symposium on Mobile Ad hoc Networking & Computing*. ACM, 2003, pp. 76–87.
[8] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," *Wireless networks*, vol. 11, no. 4, pp. 471–487, 2005.
[9] C. M. Thomas and K. M. Nielsen, "Mechanisms of, and barriers to, horizontal gene transfer between bacteria," *Nature reviews microbiology*, vol. 3, no. 9, pp. 711–721, 2005.
[10] J. Adler, "Chemotaxis in bacteria," *Science*, vol. 153, no. 737, pp. 708–716, 1966.
[11] H. C. Berg and D. A. Brown, "Chemotaxis in escherichia coli analysed by three-dimensional tracking," *Nature*, vol. 239, no. 5374, pp. 500–504, 1972.
[12] L. Fischer-Fantuzzi and M. Di Girolamo, "Triparental matings in escherichia coli," *Genetics*, vol. 46, no. 10, p. 1305, 1961.
[13] M. Achtman, G. Morelli, and S. Schwuchow, "Cell-cell interactions in conjugating escherichia coli: role of f pili and fate of mating aggregates." *J. Bacteriol.*, vol. 135, no. 3, pp. 1053–1061, 1978.
[14] D. G. Gibson, J. I. Glass, C. Lartigue, V. N. Noskov, R.-Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie *et al.*, "Creation of a bacterial cell controlled by a chemically synthesized genome," *Science*, vol. 329, no. 5987, pp. 52–56, 2010.
[15] R. P. Novick, "Plasmid incompatibility," *Microbiological reviews*, vol. 51, no. 4, p. 381, 1987.
[16] M. Couturier, F. Bex, P. Bergquist, and W. Maas, "Identification and classification of bacterial plasmids." *Microbiological reviews*, vol. 52, no. 3, p. 375, 1988.
[17] M. Holz and S.-H. Chen, "Spatio-temporal structure of migrating chemotactic band of escherichia coli. i. traveling band profile," *Biophysical Journal*, vol. 26, no. 2, pp. 243–261, 1979.
[18] E. F. Keller and L. A. Segel, "Traveling bands of chemotactic bacteria: a theoretical analysis," *J. Theo. Biol.*, vol. 30, no. 2, pp. 235–248, 1971.
[19] D. B. West, *Introduction to graph theory*. Prentice hall Upper Saddle River, 2001, vol. 2.
[20] A. Hayter, *Probability and statistics for engineers and scientists*. Cengage Learning, 2012.
[21] J.-M. Ghigo, "Natural conjugative plasmids induce bacterial biofilm development," *Nature*, vol. 412, no. 6845, pp. 442–445, 2001.
[22] T. Ahmed, T. S. Shimizu, and R. Stocker, "Microfluidics for bacterial chemotaxis," *Integrative Biology*, vol. 2, no. 11-12, pp. 604–629, 2010.

Nabiul Islam received B.Tech and M.Tech degrees in Computer Science and Engineering from University of Calcutta in 2008 and 2010, respectively.

He is currently pursuing PhD at IIT Kharagpur, India. His research interests include design and analysis of energy-efficient protocols in the area of energy constrained networks such as Nanonetworks, Wireless Body Area Networks.

Sudip Misra is an Associate Professor in the School of Information Technology at the Indian Institute of Technology Kharagpur. He received his Ph.D. degree in Computer Science from Carleton University, in Ottawa, Canada. His current research interests include algorithm design for emerging communication networks. Dr. Misra is the author of over 200 scholarly research papers. He has won eight research paper awards in different conferences. He was awarded the IEEE ComSoc Asia Pacific Outstanding Young Researcher Award at IEEE GLOBECOM 2012, Anaheim, California, USA. He was also the recipient of several academic awards and fellowships such as the Young Scientist Award (National Academy of Sciences, India), Young Systems Scientist Award (Systems Society of India), Young Engineers Award (Institution of Engineers, India), (Canadian) Governor Generals Academic Gold Medal at Carleton University, the University Outstanding Graduate Student Award in the Doctoral level at Carleton University and the National Academy of Sciences, India Swarna Jayanti Puraskar (Golden Jubilee Award). He was awarded the Canadian Governments prestigious NSERC Post Doctoral Fellowship and the Humboldt Research Fellowship in Germany. Dr. Misra is the Editor-in-Chief of the International Journal of Communication Networks and Distributed Systems (IJCNDS), Inderscience, Switzerland. He has also been serving as the Associate Editor of the Telecommunication Systems Journal (Springer SBM), Security and Communication Networks Journal (Wiley), International Journal of Communication Systems (Wiley), and the EURASIP Journal of Wireless Communications and Networking. He is also an Editor/Editorial Board Member/Editorial Review Board Member of the IET Networks and IET Wireless Sensor Systems. Dr. Misra has 8 books published by Springer, Wiley, and World Scientific. He was invited to chair several international conference/workshop programs and sessions. Dr. Misra was also invited to deliver keynote/invited lectures in over 30 international conferences in USA, Canada, Europe, Asia and Africa.