INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 13(2), 192-204, April 2018.

Decoupling 5G Network Control: Centralised Coordination and Distributed Adaptation

S.J. Lin, J.G. Yu, W. Ni, R.P. Liu

Shangjing Lin*, Jianguo Yu
Beijing University of Posts and Telecommunications
Beijing, China, 100876
linshangjing@bupt.edu.cn, yujg@bupt.edu.cn
*Corresponding author: linshangjing@bupt.edu.cn

Wei Ni
Data 61, CSIRO
Australia, 2122
Wei.Ni@data61.csiro.au

Ren Ping Liu
University of Technology, Sydney
Australia, 2122
Ren.Liu@uts.au

Abstract: Fifth generation mobile networks (5G) will be featured by miniaturised cells and massive dense deployment. Traditional centralised network control cannot adapt to high signalling delay, and is therefore not scalable for future 5G networks.To address this issue, we adopt the software-defined networking (SDN) approach of decoupled network control and data transmission. In particular, delay-sensitive interference suppression for data transmission is decoupled from delay-tolerant topology control and base station coordination. This substantially alleviates the requirement of network control on delay and complexity, hence simplifying 5G control plane design, reducing signalling overhead, and enhancing network scalability. Case studies show that our decoupled network control is effective for timely interference mitigation and reliable topology management. The stability and scalability of our approach are also demonstrated.
Keywords: fifth generation mobile network (5G); network control; protocol stack; standardisation.

1 Introduction

Fifth generation mobile networks (5G) will be featured by miniaturised cells and massive dense deployment [3]. The features are driven by the goal of a thousand times increased network capacity, compared to the current 4G systems [13]. They are further triggering a paradigm shift of network architecture and operation. Privately owned/installed plug-and-play base stations and individually subscribed backhaul connections are promising in future 5G deployments.

Fig. 1 illustrates the promising future 5G architecture, where plug-and-play 5G base stations are connected to the Servicing Gateways (S-GWs) of the 5G core network for interference coordination and mobility control. The backhaul connections between the base stations and the S-GW can be heterogeneous, subscribed to different Internet Service Providers (ISPs) (like existing household WiFi). Such architecture is able to support the massive dense deployment of 5G base stations, using existing network infrastructure and saving deployment cost.

Meanwhile, critical challenges are arising in the control of the heterogeneous 5G networks. Particularly, high and substantially unbalanced delay of hundreds of milliseconds may occur on the subscribed backhaul connections between the base stations and the S-GW [10]. Close and

timely network control that has been necessary in existing cellular systems (with dedicated transport networks) for interference mitigation [1] becomes impossible in the heterogeneous 5G systems. However, timely control will remain crucial in drastically changing wireless environments where interference levels can change instantly. Fast power control at intervals of milliseconds will still be necessary to mitigate interference in future 5G systems.

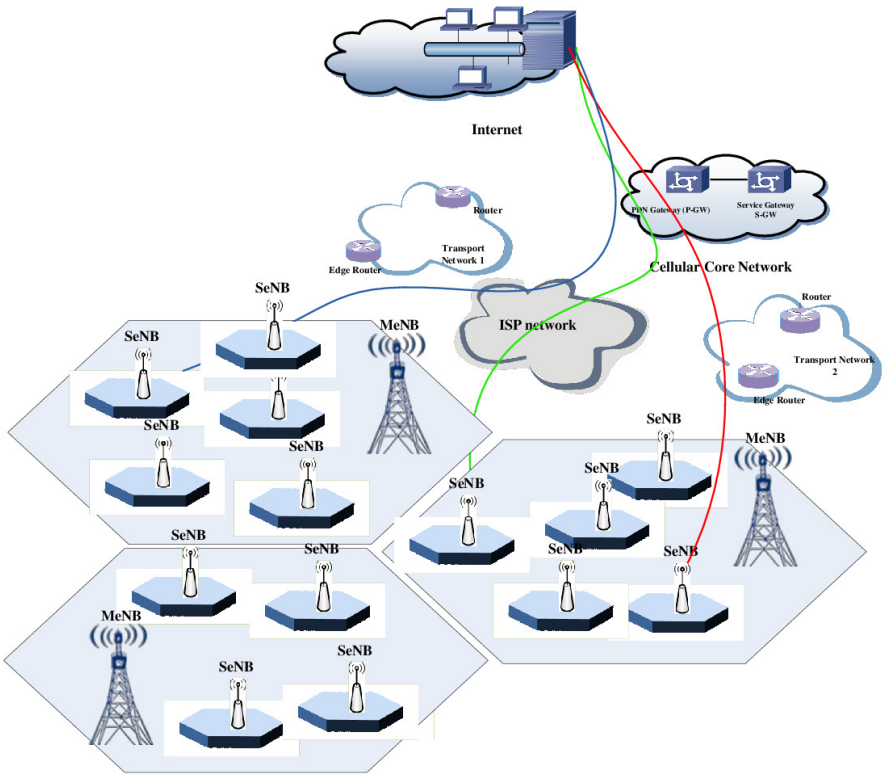

Figure 1: An illustration on the heterogeneous 5G architecture, where some base stations are connected to the S-GW via cellular infrastructure, as highlighted by the red path, while most base stations are connected to the S-GW through various ISP networks, as indicated by the green and blue paths. Centralised control signalling travels via these colour-coded paths.

Earlier, the concept "software-defined networking (SDN)" was proposed to split control and data in wired networks, which can simplify network control and reduce cost [5]. The SDN concept was experimentally demonstrated effective for its application to wired IP/Ethernet transport networks of mobile systems [11]. Recently, the SDN idea of centralised control and distributed data delivery has been extended for the control of radio access networks (RANs) [4, 6–9, 11]. Nevertheless, all the extensions would experience the above issue of delayed control if applied to the heterogeneous 5G networks, and fail to suppress interference in a timely manner.

One of the SDN extensions to wireless applications was big base station abstraction [4], where a controller models base stations as a dimension of resources, and assigns them and powers to traffic flows. Unfortunately, the abstraction was based on an oversimplified assumption of independence between base stations and frequencies. As a consequence, interference, which can substantially differ in practice when different pairs of base stations are allocated with the same frequency, was treated as static.

Other extensions of the SDN centralised network control were on software-defined real-time pairing of antennas and base stations, adapting to traffic demand and interference [7–9]. They are promising for a local 5G deployment with a dedicated (operator controlled) transport network.

However, the extensions are unsuitable for the heterogeneous 5G architecture involving ISP transport networks and incurring delayed control. They are also unsuitable for large and densely deployed systems, where centralised optimisations are computationally prohibitive and limit the network scalability.

In this article, we propose to decouple 5G network control and distribute control computations, so that the critical issues of delayed control and limited scalability can be tackled. In particular, delay-sensitive interference suppression for data transmission is decoupled from delay-tolerant topology control and base station coordination. The interference suppression can be achieved quickly through distributed adaptation of base stations. The topology and frequencies are progressively coordinated through a centralised SDN controller. As a result, the requirement of network control on delay and complexity can be substantially alleviated, simplifying 5G control plane design, reducing signalling overhead, and enhancing network scalability. Case studies show that our decoupled network control is effective for timely interference mitigation and reliable topology management. The stability and scalability of our approach are also demonstrated.

2 Evolutionary view of 5G network control

As discussed in Section 1, centralised network control faces critical challenges of severely delayed control and limited scalability in a heterogeneous 5G network depicted in Fig. 1. Decentralisation of network control is the way to eliminate the challenges, and therefore of practical value.

We note that a 5G base station will not only be a signal generation and processing entity, but it will also have self-organising capabilities of running a variety of operations specified by software. In this sense, 5G base stations will be equipped with a certain level of intelligence, and capable of automation and local decision-makings. The decisions that a base station can make can include switching from severely interfered (crowed) wireless channels to less interfered channels, as well as adjusting its own transmit power.

Given the intelligence of individual base stations, the entire network can be visualised as an evolution process driven by changing traffic demand, wireless channel conditions and interference levels. In particular, the network topology and the radio resources assigned with the topology evolve along with those changes. The network topology here describes clusters of cells in the 5G network, where cells belonging to a cluster reuse the same frequency assigned to the cluster. Each individual of the base stations drives the evolution of the topology by forming clusters and switching between the clusters (i.e., from severely interfered clusters to less interfered clusters). The entire population of the base stations evolves towards stable topology and balanced interference.

Fig. 2 shows an example of the network evolution, where only four base stations and two clusters are considered for illustration convenience. The evolution consists of two stages. In the first stage, the base stations within each cluster quickly and interactively adjust their transmit power through distributed automation to balance interference. In the second stage, the base stations choose to stay in the current clusters or switch out to less interfered clusters. This is a speciation stage in the evolution. It requires a brief idea of the entire base station population, and therefore needs centralised assistance. The two stages take place in an alternating manner, until no base stations want to switch clusters and the system is stabilised.

Such an evolutionary view of network control allows for the decomposition of the network control into centralised and distributed parts. In light of this, a decoupled, hybrid protocol stack for 5G network control is developed, as will be described in Section 3.

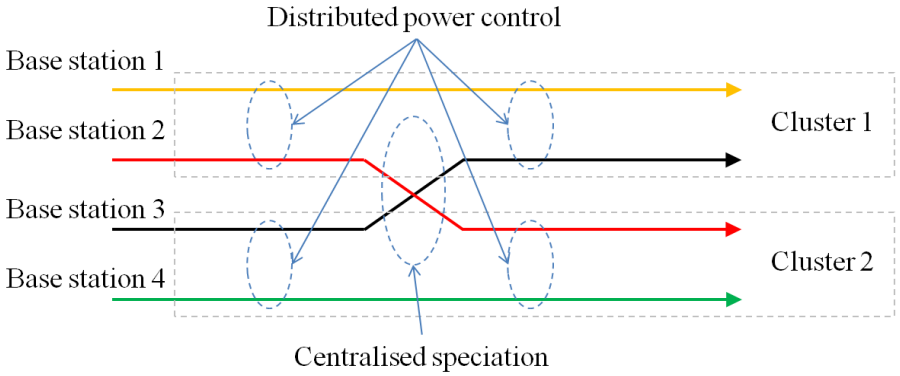

Figure 2: An illustration on the evolution modelling of 5G networks, where colour-coded arrow lines represent different base stations.

It is worth pointing out that IEEE 802.11 access point (AP) can also be treated as rational with self-organising capabilities. However, the APs cannot interact with each other, and therefore are unable to suppress interference. What they can do is to detect and select radio channels with negligible interference on the installation and initialisation.

3 Decoupled, hybrid protocol stack for 5G network control

The new protocol stack that we propose for 5G network control consists of two layers, namely, upper layer and lower layer. The two layers are designed to decouple the delay-tolerant and delay-sensitive aspects of network control, driven by the evolutionary view of 5G network established in the last section.

a. The upper layer of the protocol stack is a centralised, delay-tolerant network control, which resides in both base stations and the network controller, tolerates hundreds of milliseconds delay over transport networks, and takes the responsibility of slow-changing topology control and resource management.

b. The lower layer of the protocol stack is a distributed, delay-sensitive network control, which resides in the base stations, and is implemented by the automation of 5G devices to combat fast-changing wireless channels and interference at an interval of milliseconds.

The two layers are interactive, distributing computations and balancing the complexity. As a result, signalling overhead can be reduced and network scalability can be enhanced.

3.1 Upper layer: Centralised coordination

The upper layer of the new 5G protocol stack is to manage the network topology and frequency, so as to balance interference and satisfy traffic demands. The upper layer is interconnected in a star topology from the centralised network controller to individual base stations.

The network topology can be described by the clusters of the cells in the network, denoted by $\mathcal{C}_n$ $(n = 1, \cdots, N)$. N is the number of clusters. Base stations belonging to each of the clusters reuse the same frequency. The base stations may need to switching between clusters (i.e., from the severely interfered clusters to the less interfered) to alleviate interference and meet traffic demand. This causes the dynamics of the network.

The upper layer of the new protocol stack models the network dynamics as an evolutionary (game) process. The recent SDN concept can be extended as such that the centralised network

controller provides guidance to the base stations, assisting them to make local cluster-switching decisions. The network controller can also steer the network evolution towards stabilised topology and balanced traffic. The reason for allowing local (distributed) decision-makings is that the globally optimal cluster-switching decisions are computationally and practically impossible in densely deployed 5G small cells, as discussed in Section 1.

The centralised network controller can be designed to distribute the available 5G spectrum between the clusters $\mathcal{C}_1, \cdots, \mathcal{C}_N$ in such a fair fashion that the total achievable data rate of each cluster is equally proportional to the request. Methods, such as Shapely value of cooperative games [12], can be used to distribute the frequencies.

The centralised controller can then evaluate the current clustering by measuring the probability distribution of the achievable data rate r_i of every cell i in a cluster $\mathcal{C}_n$ ($i \in \mathcal{C}_n$). The distribution function is denoted by F_n. The controller further evaluates the probability distribution with regards to all the clusters. The distribution function is denoted by F.

For the clusters whose distribution functions F_n are left to F (i.e., the clusters have lower achievable data rates), the centralised network controller specifies probabilities for their base stations to switch to other less interfered clusters whose distribution functions are right to F.

The probabilities can be specified, depending on the ranking of the base stations within their current clusters and the distances between the distributions of the current clusters and the target clusters. The base stations will have higher probabilities to switch out of a cluster $\mathcal{C}_i$, if F_i is further away from left to F (i.e., $\mathcal{C}_i$ is crowded). They will have higher probabilities to switch into $\mathcal{C}_j$, if F_j is further away from right to F. $i \neq j$. Within the cluster $\mathcal{C}_i$, the lower data rate that a base station can achieve, the probability will be higher for the base station to switch out of the cluster, since the base station suffers more severe interference under the current clustering.

The first step of the clustering review is to specify the ratio of the femtocells that need to switch clusters. This is an evolutionary process. The ratio of the femtocells that need to switch out of $\mathcal{D}_m$ can be specified by

$$\begin{aligned}\dot{p}_m =& p_m \times \|f_m(x) - \bar{f}(x)\|_{\mathcal{D}} \\ =& \frac{p_m}{\sum\limits_{\forall m} Q_m} \text{sign}\Big(\mathbb{E}_m(x) - \mathbb{E}(x)\Big) \times \\ & \underbrace{\left(1 - \int_0^{\infty} \min\Big(f_m(x), \bar{f}(x)\Big) dx\right)}_{Q_m},\end{aligned} \tag{1}$$

where p_m is the ratio of femtocells in $\mathcal{D}_m$; $\|\cdot\|_{\mathcal{D}}$ denotes the statistical difference between two clusters, which can be calculated by $Q_m / \sum_{\forall m} Q_m$; sign$(\cdot)$ is the sign function.

Q_m denotes the under-braced part, in which the integration calculates the area of $f_m(\cdot)$ that is not overlapped with $f(\cdot)$. $\mathbb{E}_m(\cdot)$ and $\mathbb{E}(\cdot)$ are the means associated with the two PDFs. The sign function indicates that the calculated non-overlapping area is to the left of $f(\cdot)$, if the sign is negative; or to the right of $f(\cdot)$, otherwise. Q_m is normalized by $1/\sum_{\forall m} Q_m$.

Such design of $\dot{p}_m$ allows the evolution process of clustering towards reducing the distributional differences between the clusters. Clusters that are below average (i.e., their PDFs are to the left of $f(\cdot)$) will have their PDFs adjusted rightwards by having some of their femtocells switch to other clusters. Specifically, $\dot{p}_m < 0$ means that $-\dot{p}_m$ of $|\mathcal{D}_m|$ femtocells should switch out of $\mathcal{D}_m$, and $\dot{p}_m > 0$ means that $\dot{p}_m$ of $|\mathcal{D}_m|$ femtocells should switch into $\mathcal{D}_m$. For a femtocell that is to switch out of $\mathcal{D}_m$ ($\dot{p}_m < 0$), it switches into $\mathcal{D}_j$ at the probability of $\dot{p}_j / \sum_{\forall i:\dot{p}_i > 0} \bar{p}_i$ ($j \neq m$).

Next, the second step of the clustering review and adjustment is for every femtocell $i \in \mathcal{D}_m$ to specify its own probability to switch clusters, denoted by $\dot{p}_{m,i}$. This is based on $\dot{p}_m$ and the

satisfaction level of femtocell i. This step differentiates the femtocells within the same cluster in terms of switching probability.

To do this, we develop the following design.

$$\dot{p}_{m,i} = \dot{p}_m \frac{\lambda_m e^{\lambda_m}}{e^{\lambda_m} - 1} \exp\Big(- \lambda_m F_m\big(x_i\big)\Big) \tag{2}$$

where $x_i = R_i^{\text{gnt}}/R_i^{\text{req}}$, and $F_m(\cdot)$ is the cumulative distribution function (CDF) of the satisfaction levels within $\mathcal{D}_m$. λ_m is a parameter that can be adjusted so that $\dot{p}_{m,i} \leq 1$. The granted data rate $R_i^{\text{gnt}} = \phi_m \omega_c \log_2(1 + \gamma_i^*)$, $i \in \mathcal{D}_m$.

The rationale for this design is that (2) does not change the overall clusterwise switching probability of the cluster, i.e., $\dot{p}_m$. Specifically, we have

$$\begin{aligned} \lim_{|\mathcal{D}_m| \to \infty} \frac{1}{|\mathcal{D}_m|} \sum_{i \in \mathcal{D}_m} \dot{p}_{m,i} &= \int_0^\infty \dot{p}_{m,i} f_m(x) dx \\ &= \dot{p}_m \frac{\lambda e^{\lambda_m}}{e^{\lambda_m} - 1} \int_0^\infty \exp\Big(- \lambda_m F_m(x)\Big) f_m(x) dx \\ &= \dot{p}_m \frac{e^{\lambda_m}}{e^{\lambda_m} - 1} \int_0^{\lambda_m} e^{-u} du = \dot{p}_m. \end{aligned} \tag{3}$$

Meanwhile, individual femtocells in the cluster can have different switching probabilities, depending on their individual satisfaction levels. As such, femtocells that receive excessive interference in crowded clusters get high possibilities to switch to other less interfering clusters.

Note that the second step of creating differential switching probabilities in a cluster is new. Individual femtocells' statuses are leveraged without invalidating the evolution of the entire population (which was specified in the first step). In contrast, conventional evolutionary games are focused on the entire population, ignoring the differences between individuals.

As an effect of the two-step design, the satisfaction distributions of different clusters converge through the evolution of clustering. The average satisfaction level of all the femtocells gradually grows until the convergence.

After the second step, each femtocell switches its cluster based on the probability specified in (2). The new M clusters, $\{\mathcal{D}_1, \cdots, \mathcal{D}_M\}$, will be considered in the next round of distributed power adaptation (as described in Section 3.2).

3.2 Lower layer: Distributed adaptation

The lower layer of the new 5G protocol stack describes a distributed automation process of every base station in terms of transmit parameters, adapting to the network topology change driven by the upper layer. Specifically, every base station, say base station i in the nth cluster ($i \in \mathcal{C}_n$), collects interference measurement reported by its mobile terminal and independently updates its transmit power P_i. The base station does so iteratively, until the transmit power is stabilised. The cluster also becomes stabilised.

A well designed utility function, with which the transmit power is independently updated in response to the measured interference, is the key to the lower layer of the protocol stack. The function will lead individual base stations to stabilise and converge in a distributed manner to the stable point of an entire cluster. This requires the function to have a convex/concave structure with respect to all the transmit powers of the cluster. Meanwhile, the data rate of each base station should not be compromised at the stable point for the convexity/concavity.

An example of such function is given by

$$U_i\left(P_i, \mathbf{P}_{-i}\right) = \arctan\Big(\frac{\gamma_i\left(P_i, \mathbf{P}_{-i}\right)}{\Gamma_i} \alpha_n\Big) - \theta_n P_i, \ i \in \mathcal{C}_n \tag{4}$$

where $\mathbf{P}_{-i}$ collects P_j for $j \in \mathcal{C}_n$ and $j \neq i$; $\gamma_i(P_i, \mathbf{P}_{-i})$ is the signal-to-interference-plus-noise ratio (SINR) of the terminal served by cell i; Γ_i is the target SINR that is required to meet the traffic demand of the terminal; α_n is a predefined coefficient to adjust the speed of the stabilisation of P_i in cluster $\mathcal{C}_n$; θ_n is a predefined coefficient to adjust the weights of the two terms in the function.

$U_i(P_i, \mathbf{P}_{-i})$ is designed to be maximized at each base station $i \in \mathcal{C}_n$. The first term at the right-hand side of the function defines the utility which drives the achieve SINR towards the target. This is through increasing the transmit power P_i. The second term at the right-hand side of the function is the cost/penalty to achieve the utility. It keeps the growth of P_i under control and prevents generating excessive interference to other cells.

The function can ensure the stabilisation of each transmit power, e.g., P_i. This is due to the concave structure of the function with respect to all the P_k's with $k \in \mathcal{C}_n$ [2]. The concavity can be rigorously proved by evaluating the Hessian matrix of the function and confirming that the Hessian matrix is negative definite with carefully selected α_n and θ_n. As a result, the function has a unique stable point that is the global optimum of the function, as given by

$$P_i^* = \min\left(\frac{I_{-i}(\mathbf{P}_{-i})\Gamma_i}{\alpha_n g_i}\sqrt{\frac{1-\theta_n}{\theta_n}}, P_{\max}\right), \tag{5}$$

where $I_{-i}(\mathbf{P}_{-i})$ is the total interference from other cells in cluster $\mathcal{C}_n$ to cell i, g_i is the channel gain of base station i to its terminal, and $P_{\max}$ is the maximum transmit power of the base station. All the cells will be stabilised at the point.

The concavity is also crucial to ensure all the cells to be stabilised at the stable point through distributed automation [12]. This is due to the fact that, on each plane cutting through the concave structure, the intersecting curve is concave. In other words, $U_i(P_i, \mathbf{P}_{-i})$ is concave with respect to P_i for any given $\mathbf{P}_{-i}$. The base station i is able to independently update the transmit power P_i towards increasing the utility, based on measured interference.

Let $\tilde{I}_{-i}(k)$ denote the k-th measurement result of interference ($k = 1, 2, \cdots$). It can be measured by the mobile terminal on the air interface, and reported to the base station through the virtual interface "I_d". $\tilde{I}_{-i}(k)$ can replace $I_{-i}(\mathbf{P}_{-i})$ in (5) and update $P_i^*(k+1)$ in practical distributed implementations, pushing all the base stations to converge to and stay at the unique stable point of (5). As a result, the requirement of signalling exchange between the cells can be eliminated, as well as the centralised coordination of the upper layer of the protocol stack.

Other utility functions with convex/concave structures can also be designed. Stable points with increased convergent SINRs are expected with proper designs of the functions.

3.3 Interactions between the two layers

The two layers of the new 5G protocol stack are interactive, as illustrated in Fig. 3. On one hand, the distributed automation of the lower layer depends on the network topology and resource management outcome of the centralised control of the upper layer. On the other hand, the centralised network control of the upper layer needs to operate on stabilised systems after the distributed automation of the lower layer, so that the network topology and resources can be adapted and updated.

The interaction between the two layers is facilitated by the interface "I_c/d" at each base station. In the downstream direction, the evolving topology that was updated at the upper layer and carried by "I_c" is passed down to the lower layer through "I_c/d" to activate distributed power control at the lower layer. In the reverse direction, transmit powers stabilised at the lower layer are handed over to the upper layer through "I_c/d", and will gather at the centralised network controller through "I_c" to trigger a new round of topology management.

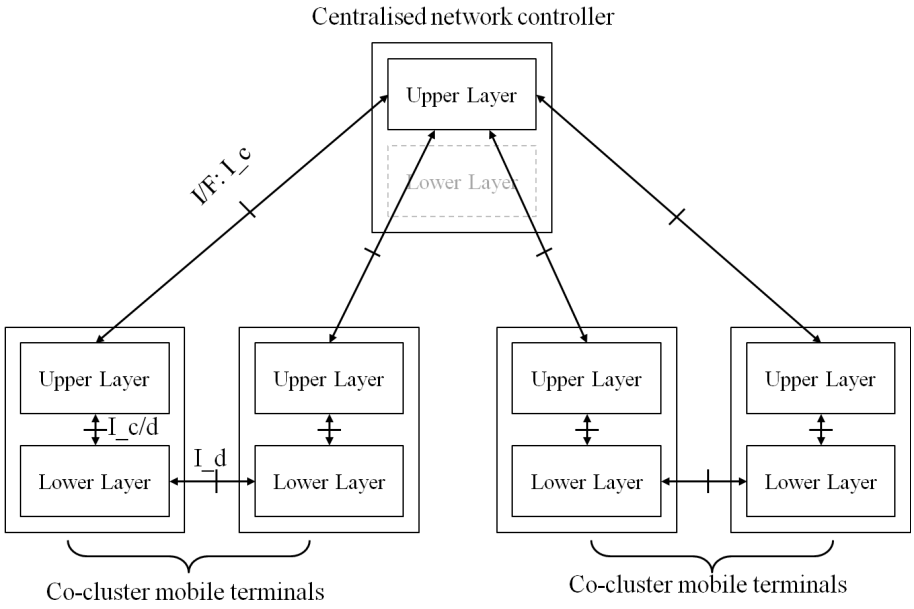

Figure 3: An illustration on the decoupled 5G control protocol stack, where the upper layer of the protocol stack is interconnected in a centralised manner through the interface (I/F) "I_c" between base stations and the centralised network controller, the lower layer of the protocol stack behaves in a distributed manner through a virtual interface "I_d", and the two layers are interactive through the interface "I_c/d" with every base station

The star topology of the upper layer also provides signalling bearers via the interface "I_c" to carry the probabilities to the base stations. On the receipt of the probabilities through the upper layer, each base station independently decides whether to switch clusters and which cluster to switch to.

As a result of the centralised upper layer, cells migrate from crowded clusters to less interfered clusters. The distributions of data rate converge across the clusters, as well as getting less spanned. Fairness is improved between the clusters, and also within a cluster. Moreover, the evolution design of the upper layer eliminates the requirement of signalling exchange between clusters, and tolerates the great level of uncertainty in large practical wireless networks. The design also distributes the complexity of clustering decision-makings between the centralised controller and individual base stations. The scalability of the network is therefore enhanced.

4 Performance studies

Extensive MATLAB simulations have been conducted to evaluate the hybrid protocol stack for 5G network control. The case where the evaluation is carried out is a heterogeneous 5G network covering a geographical area of $500^2\pi \mathrm{m}^2$, where there are 30 available frequency channels and each channel has a bandwidth of 180 KHz.

Consider that the privately owned/installed plug-and-play base stations are most likely to be indoors and operate at a low power level. We assume that the maximum transmit power of each base station is 20 dBm. We also assume a wall penetration loss of 10 dB when the radio signals of a 5G cell leak out from indoors to outdoors and become interference to other cells. The path loss exponent is 3.7. The receive noise at the mobile terminals is set to be -174 dbm/Hz. The traffic demand is assumed to be 5.0 Mb/s per mobile terminal.

Fig. 4 shows the overall throughput of the 5G network with the increasing number of cells, where the number of clusters grows from one to five. Apart from our hybrid approach, we also plot the pure centralised SDN network control proposed in [6], where clusters are constructed using graph theory to maximize the distance between the clusters, and the base stations transmit a fixed power of 20 dBm since fast power control at milliseconds intervals is impossible due to centralised and severely delayed control. The figure demonstrates that, in general, the hybrid network control provides higher throughput than the pure centralised control. When the evolutionary clustering is on, the hybrid network control adaptively decides the number of clusters. The throughput under the hybrid network control operates as the topmost of all the five solid curves, which is far beyond the throughput of the centralised approach.

A close look at each of the five solid curves shows that for every given number of clusters, the throughput starts by growing, due to the increased traffic demand resulting from the increased cell density. After reaching a peak, the throughput rapidly drops. The reason is because the network is so dense, the clusters are becoming very crowded, and the interference becomes excessively large. Distributed automation of power control or even switching between clusters cannot help to suppress the interference. The excessive interference not only offsets the increased throughput resulting from increased demand, but also adversely affects and further reduces the throughput. In this sense, the peak corresponds to the case where the network density and interference are balanced and the throughput is maximized given the number of clusters.

A joint look across all the five solid curves in Fig. 4 shows that increasing the cluster number is able to increase the peak throughput, as well as the number of cells achieving the peak. Of course, clustering can help alleviate interference by reducing the number of cells using the same frequency, as shown in the figure where the number of cells that can be supported increases with the number of clusters. On the other hand, clustering also decreases the bandwidth that every cell can use, as depicted in the figure where the throughput decreases with the increasing number of clusters in the presence of a small number of cells (i.e., $N \leq 50$). The conclusion that we draw is that the significantly reduced interference due to the evolutionary clustering is able to compensate for the reduced bandwidth, when the cell density and interference are balanced. As a result of this, the peak throughput keeps growing with the cluster number.

Fig. 5 plots the average transmit power required to achieve the throughput performance in the previous figure, with respect to the increasing number of cells. The transmit power is the average of all the base stations' after they are stabilised from the distributed power control automation. We can see that, in general, the stabilised transmit power decreases with the increase of the cluster numbers. This is because clustering helps alleviate interference, which in turn reduces the requirement of transmit power. When the evolutionary clustering of our hybrid network control is on, the average transmit power is as shown as the highlighted black zigzag curve in the figure. This is because the number of clusters is adaptively adjusted to maximize the throughput, when the evolutionary clustering is on. The number of clusters may increase, as the number of active cells grows. This can lead to a sudden drop of the transmit power, partly because of the alleviated interference and partly because of the reduced bandwidth per cell.

It is also interesting to notice that the peaks on the highlighted black curve grows slowly with the number of cells. This is consistent with Fig. 4, indicating that increased transmit powers are necessary to improve the achievable throughput. The sudden drop of the transmit power from the peaks also indicates that when interference becomes intensive, releasing part of the frequencies for a new cluster can substantially relieve the interference. As a result, much lower transmit powers are required to achieve the same throughput. In other words, the peaks are where the base stations should trade out their bandwidth for relieved interference.

Fig. 6 shows the convergence of the average throughput of each cluster along with evolution over time, where a single evolution process of 5 clusters is recorded. It demonstrates the conver-

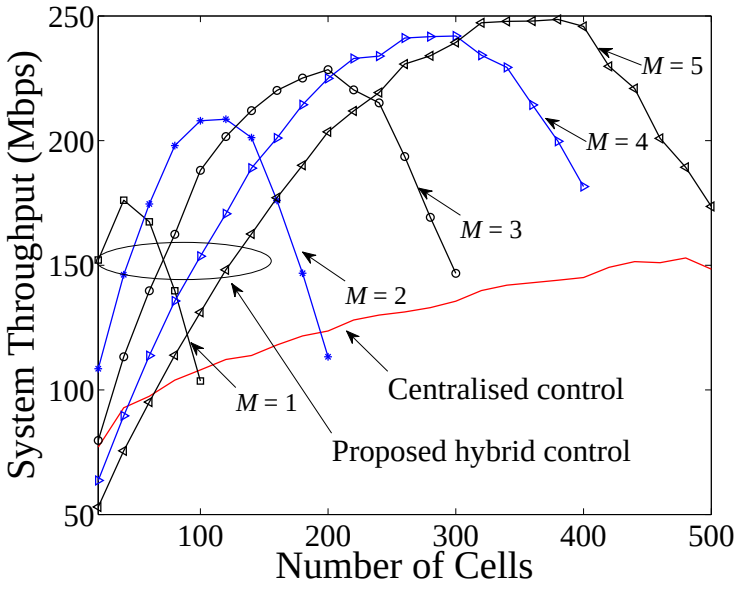

Figure 4: Overall system throughput versus the number of cells, where the number of clusters, denoted by N, grows from 1 to 5.

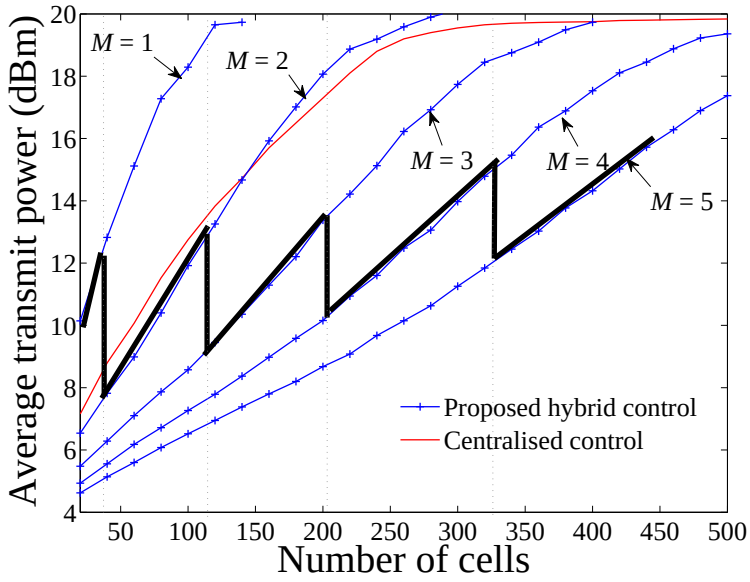

Figure 5: Average transmit power per base station versus the number of cells, where the black curve is the actual required average transmit power when the evolutionary clustering of our hybrid network control is on, and is plotted by picking the crossing points of any two adjacent throughput curves in Fig. 4.

gence of the clusters in terms of throughput. The largest difference among the clusters decreases from 0.8 Mbps to 0.1 Mbps after 320 evolution stages (i.e., 6.4 seconds), and the difference diminishes by 500 evolution stages (i.e., 10 seconds). We also see that most of the clusters can

fast converge to balancing their throughput. Specifically, clusters 1 to 4 converge by 200 stages. Cluster 5 does not converge as fast as the others do, because of its small initial size (as will be shown in Fig. 7). However, the convergent tendency of the cluster is clear, as shown in the figure.

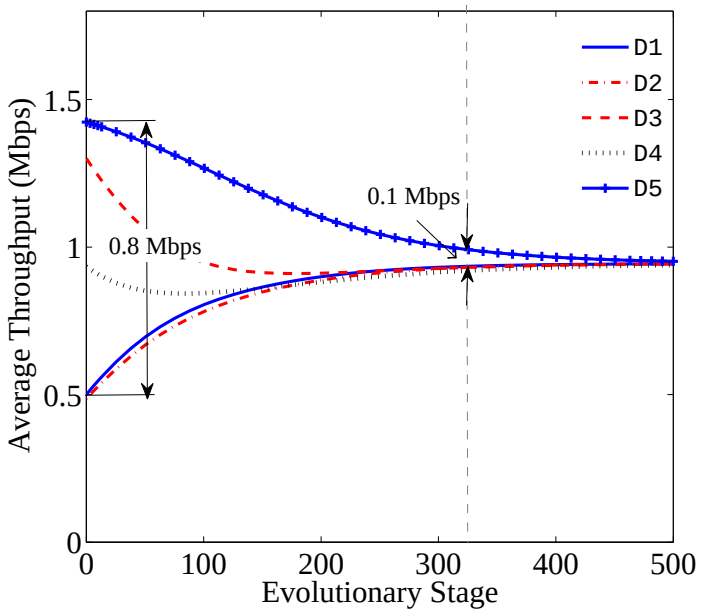

Figure 6: Convergence of the throughput per cell with the growth of evolution stages, where $N = 150$ and $M = 5$.

Fig. 7 shows the stabilization process of the size of each cluster along with evolution over time, corresponding to Fig. 6. We can see the difference of the cluster sizes decreases fast with the evolutionary clustering. However, the cluster sizes do not converge to the same size, as opposed to the throughput curves of the clusters which evolve to converge, as shown in Fig. 6. The reason for this lies in the network geometry of this particular simulation, where the femtocells are geographically unevenly distributed. As a result, the cluster sizes vary so as to balance interference in each cluster.

Another interesting finding in Fig. 7 is that after the cluster sizes stabilize, the throughput of the clusters continues converging. Specifically, by 320 evolution stages, the cluster sizes start to stabilize, while the maximum throughput difference is still about 0.1 Mbps. After that, local adjustment of clustering is happening without significantly changing the cluster sizes. The clustering evolution enables the clusters to locally swap their femtocells (e.g., between clusters 4 and 5 in Fig. 7) and adjust their geometry, until each of the clusters is most widely spanned in space and intra-cluster interference is alleviated.

Figs. 6 and 7 also confirm the effectiveness of our proposed framework, as they evidence that the convergence of throughput among the clusters unnecessarily requires the convergence of the sizes of the clusters. Our proposed framework, especially the cluster conformation design of (1), targets to balance the throughput (or in other words, balance the interference), rather than the cluster sizes.

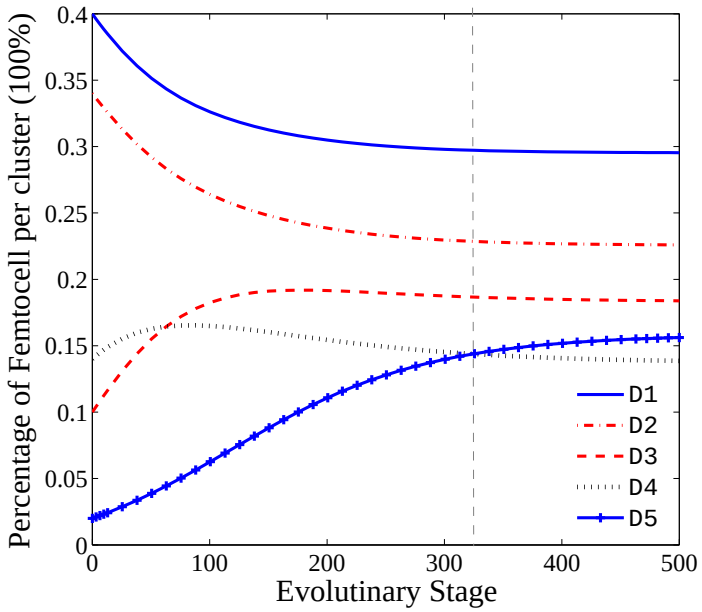

Figure 7: Convergence of the cluster size per cluster with the growth of evolution stages, where $N = 150$ and $M = 5$.

5 Prospective standardisations

Given the superior performance, the new hybrid network control is promising for future 5G systems. Standardisations will be important, focusing on the three interfaces of I_c, I_d, and I_c/d. The cycle that the centralised speciation takes place needs to be specified, which has strong impact on the convergence speed of the network evolution, as described in Section 3.1.

Standardisations will also be required on the utility functions described in Section 3.2, which are critical for the convergence of the distributed power control automation, as well as the effectiveness of interference suppression.

Other standardisation efforts will be on identifying typical probability distribution functions for the centralised controller to describe the distribution of data rate in each cluster, as described in Section 3.1. This will be important to leverage the complexity of the centralised SDN controller, as well as the reliability of the network evolution.

6 Conclusions

In this article, we proposed to decouple 5G network control into two layers, namely, centralised coordination at the SDN controller and distributed adaptation of base stations. Delay-tolerant topology control and delay-sensitive interference suppression are respectively handled by the two layers. The decoupling was based on a new evolutionary view of 5G network control. The decoupled network control is able to alleviate the requirement of delay and complexity, and therefore enhance network scalability. Case studies show that our decoupled network control is effective for timely interference mitigation and reliable topology management. The stability and scalability of our approach are also demonstrated.

Acknowledgment

This work was partly supported by National Natural Science Foundation of China under the Grant No.61701034 and No.61531007 and partly supported by China Postdoctoral Science Foundation under the Grant No 2017M620696.

Bibliography

[1] Alliance, N. G. M. N. (2012); Small cell backhaul requirements, *white paper*, June, 2012.

[2] Bemporad, A.; Heemels, M.; Johansson, M. (2010); *Networked control systems*, 406, Berlin: Springer, 2010.

[3] DOCOMO (2014); DOCOMO 5G White Paper: 5G Radio Access: Requirements, *Concept and Technologies*, 2014.

[4] Gudipati, A.; Perry, D.; Li, L. E.; Katti, S. (2013); SoftRAN: Software defined radio access network, *Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking*, 25-30, 2013.

[5] Kim, H.; Feamster, N. (2013); Improving network management with software defined networking. *IEEE Communications Magazine*, 51(2), 114-119, 2013.

[6] Lin, S.; Tian, H. (2013); Clustering based interference management for QoS guarantees in OFDMA femtocell, *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*, 649-654, 2013.

[7] Ni, W.; Collings, I. B. (2012); Indoor wireless networks of the future: adaptive network architecture, *IEEE Communications Magazine*, 50(3), 130-137, 2012.

[8] Ni, W.; Collings, I. B. (2013). A new adaptive small-cell architecture, *IEEE Journal on Selected Areas in Communications*, 31(5), 829-839, 2013.

[9] Ni, W.; Collings, I. B.; Lipman, J.; Wang, X.; Tao, M.; Abolhasan, M. (2015); Graph theory and its applications to future network planning: Software-defined online small cell management, *IEEE Wireless Communications*, 22(1), 52-60, 2015.

[10] Ni, W.; Liu, R.P. ; Collings, I.B.; Wang, X. (2013), Indoor cooperative small cells over Ethernet, *Communication Magazine, IEEE*, 51(9), 100-107, 2013.

[11] Pentikousis, K.; Wang, Y.; Hu, W. (2013); Mobileflow: Toward software-defined mobile networks, *IEEE Communications magazine*, 51(7), 44-53, 2013.

[12] Shapley, L. S. (1971); Cores of convex games, *International journal of game theory*, 1(1), 11-26, 1971.

[13] Wen, T.; Zhu, P. (2013); 5G: A technology vision, *Huawei*, 2013.