Digital Business Ecosystem

Contract n° 507953

Workpackage 19:

Peer-to-peer network behaviour

Deliverable D19.5

Report on Distributed Mechanism

Information Society Technologies

Project funded by the European Community under the "Information Society Technology" Programme

Contract Number: 507953

Project Acronym: DBE

Title: Digital Business Ecosystem

Deliverable N°:

Due dates: 11/2006

Delivery Date: 12/2006

Short Description:

This deliverable is a report on distributed mechanisms and policies to obtain different network topologies. The work presented here consists of two parts: firstly, what topology would FADA network evolve to as system requirements are changing and distributed mechanisms applied on the network vary. This part of work was done by combining graph-theoretic and system requirements. Secondly, how the interactions between the SMEs and FADA communication infrastructure affect the topology of the FADA network.

Author: HWU

Partners contributed: SUN, UBHAM

Made available to: Public

Versioning		
Version	**Date**	**Author, Organisation**
1	18/12/2006	Juan Wang and Philippe De Wilde, HWU

Quality check:

1st Internal Reviewer :

2nd Internal Reviewer:

Contents

List of Figures

List of Tables

Chapter 1

Introduction and relationship to the project

This deliverable is the final report for sub-task S7 (DBE Topology) and S8 (Multiple Scales), which are a part of work in W19, a workpackage looking at the specifics of the Peer-to-Peer behaviour in the DBE. This deliverable is a report on distributed mechanisms and policies to obtain different network topologies, which builds on some work reported in D19.2 and D19.3. Studies on how the performance of FADA depends on network topology presented in D19.2 and how the DBE network copes with communications generated by the evolutionary environment reported in D19.3, together offer insights on the issue of distributed mechanisms for different network topologies in the DBE.

The results in this study may provide suggestions to task C51 (Peer-to-peer simulator, as reported in [26]). In addition, these two sub-task interact with W11 (Dynamics of networks, reported in [37] [38] [39]), in which the study of the mathematical and empirical analysis of information propagation on different types of graphs and under different network topologies, and study on the effects of different topologies on the diversity and fitness of the system, are presented.

Chapter 2

Background and related work

Currently, the requirements on the way how we communicate and how efficient we can communicate are increasing. Computer networks are absolutely necessary in the daily life. As a consequence, network management is regarded as an important issue. In order to monitor and manage devices on the network, knowledge of the network topology is requisite.

2.1 What is topology?

Topology refers to the shape of a network, or the network's layout of connected devices on a network. It determines how different nodes in a network are connected to each other and how they communicate. Topologies are either physical (real) or logical (virtual). The way that the devices are connected to the network through the actual cables that transmit data, representing the physical structure of the network, is called the physical topology. In contrast, the logical topology is the way that the data passes through the network from one device to the next without regard to the physical interconnection of the devices. Logical topologies are bound to the network protocols that direct how the data moves across a network. A network's logical topology is not necessarily the same as its physical topology. In this report, only the logical topology is discussed. In the real world, numerous systems that are highly technological and intellectual can be described by

models of complex networks. These networks are structures consisting of nodes connected by links. For example, the Internet is a complex network of routers and computers linked by physical or wireless links; the World Wide Web (WWW) is a network of Web pages connected by hyperlinks. In this report, the SME network is a complex network of companies and the communication network is a network of FADA nodes.

Due to the complexity of interactions between nodes and their dynamic behaviours, the topologies of these networks are not possible to be categorized into one of the five basic types of network topologies: Bus, ring, star, tree and Mesh. We call these networks as complex networks.

Over the past few years, the availability of super computers and the digitization of data acquisition in all fields have made huge database available to study the topology of various real networks. At the same time, the mechanisms that determine the topology of complex networks and the interrelation between structures of networks and network dynamic behaviours have encountered significant challenges as well. For instance, how does a disease spread over social networks [23]? How do cascading failures propagate through a large power transmission grid or a global financial network [35]? These kinds of problems exist in our daily life.

These issues above can be generally concluded into answering one question, which is how the function of a network is affected by its structure or topology. For example, the spread of information and disease is affected by the topology of social and organizational networks, and the robustness and stability of power transmission is affected by the topology of the power grid. Therefore, it is important to find out how networks grow, how the pattern of connections, or the shape, is influenced by the growth process, and what mechanisms lead to each type of topology.

The structure of various networks, including social and computer networks, has been widely studied in the physics literature [8, 14]. There have been findings suggesting

that some behaviours and properties can be found in a number of different types of networks. Therefore, there are an increasing number of researchers who believe that there must be some laws describing these networks, and once these laws are found for a certain network, they can be utilized in certain other networks and explain the behaviour more correctly.

2.2 Topology of Networks

2.2.1 Topological Properties

The structure of networks has been studied by graph theory. A network is regarded as a graph, where vertices (or nodes) represent entities in the network and arcs (or links) represent interactions between the entities. Many quantities and measures of complex networks have been proposed and investigated. In this section, the most important graph-theoretic properties that are used to measure a network's topology in this report are introduced. These are degree distribution, average path length, diameter, and clustering coefficient.

Degree and Degree Distribution

The number of edges connected to a vertex is called the degree of the vertex. In a network, vertices always have different degree. The degree of a node is characterized by a distribution function P(k), which is the probability that a randomly chosen node has exactly k connections. Another important characteristic of a graph is the average degree (or average number of links), as it gives a relationship between the size of the graph and its sparseness [41]. A high average degree means individual vertices will have a very large number of neighbours, which place a higher burden on communication.

Average Path Length and Diameter

The average path length is the average of the shortest path lengths from each node to every other node. It characterizes the spread of a graph. For example, it affects the number

of "hops" that a packet must take to get from one computer to another computer on the Internet, how fast requests are propagated through a network, the time it takes for a disease to spread through a population, and so on. Another important way to characterize the spread of a graph is to calculate the diameter, which is defined as the longest distance between any two nodes.

Clustering Coefficient

A common property of social networks is the form of cliques, which represent circles of friends, where every member knows every other member. In a graph, it is called a cluster, when there is a subset of nodes that are fully connected to each other. It is quantified by the clustering coefficient [42] as the probability that two of a particular node's nearest neighbours are connected. A graph with a high clustering coefficient implies a large number of clusters, while a low clustering coefficient implies disconnected vertices. In this report, three recently most popular topologies are used as modeling paradigms: random graphs, small-world models, and scale-free models.

2.2.2 Random Graph

Network topologies such as chains, grids, lattices and fully-connected graphs are regular graphs. These topologies are simple for analysis without being bothered by complexity in the network structure itself. However, networks with a complex topology and unknown organizing principles are often random. Therefore, people start to think about graphs that are completely random.

The theory of random graphs was introduced by Erdös Rènyi (ER), as an early effort to model complex networks in [16] about 40 years ago. In an ER random graph, edges are distributed randomly and the presence or absence of any edge between two nodes in the network is dependent on a fixed connection probability p. Random graph theory aims to determine at what p, a particular property of a graph will most likely appear.

The construction (or the evolution process) of a random graph is as the following: starting with N isolated nodes, then connect every pair of nodes with probability p. The graphs obtained at different stage correspond to larger and larger p, until obtaining a fully connected graph for $p \to 1$. An example of the evolution process is shown in Fig. 2.1.

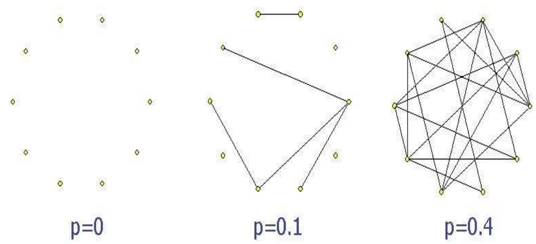


Figure 2.1: The graph evolution process for Erdös Rènyi (ER) model.

The degree distribution of a random graph has been found to follow a Poisson distribution:

$$P(k) = \binom{n}{k} p^k (1-p)^{N-k} \simeq \frac{z^k e^{-z}}{k!}, \quad k \geq 0.$$

Here k represents nodes having exactly k edges, n is the total number of edges and N is the number of nodes in the network. Moreover, the average path length of a random graph has been found to scale with the number of nodes:

$$\ell \sim \frac{\ln N}{\ln \langle k \rangle}. \tag{2.1}$$

Various types of networks in the real world have been modelled as random graphs, particularly in epidemiology studies

2.2.3 Small World Networks

Many real-world networks such as the World Wide Web, the Internet, communication networks, electric power grid and networks of movie actors, have been found to display the so-called "small-world network" phenomenon. The small-world networks are characterized by two primary properties, which are small average path length and unusually large clustering coefficient. The most famous manifestation of the small worlds is the concept of "six degree separation", which has been concluded by the psychologist Stanley Milgram in 1967 [27] that two random US citizens were connected by an average of six acquaintances.

In 1998, Watts and Strogatz firstly proposed a simple model (WS small-world model) to generate small-world networks [42]. The inspiration of the WS model is based on the fact that many networks actually lie somewhere between the extremes of order and randomness. Although regular networks and random graphs are both useful idealizations, most of the real-world networks are neither entirely regular nor entirely random. The procedure of the WS model is as the following: it starts with a ring lattice with n vertices and each vertex connects to its k nearest neighbors. Each edge is rewired at random with probability p. The parameter p is able to control the graph to be between completely regular ($p = 0$) and completely random ($p = 1$), as illustrated in Fig. 2.2.

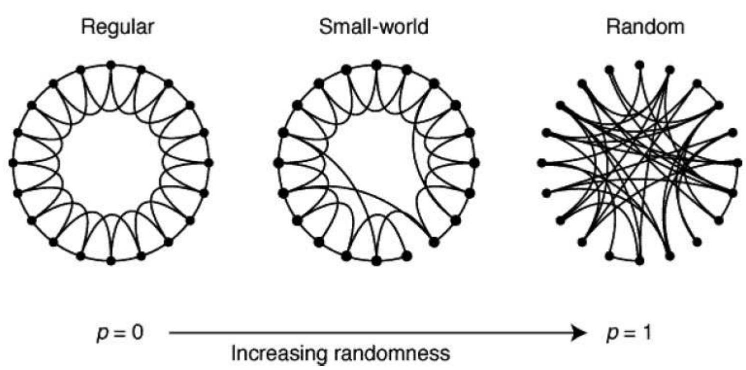


Figure 2.2: [Courtesy of [42]]The rewiring procedure of WS small-world network model.

Small-world networks model has been widely used for real-world applications, such as

oscillator networks [24], disease propagation [44], information propagation [46] and neural networks [29].

2.2.4 Scale-free Networks

A common characteristic of the ER random graph and the WS small-world model is the degree distribution of the network is homogenous, with peak at an average value. However, recently, some real-world networks, such as networks of movie actor collaboration [10, 42], science collaboration [31], WWW [6, 20] and Internet [17], have been found to have a degree distribution that follows a power-law with different exponents: $P(k) \sim k^{-r}$, which deviates from poisson distribution of the random graph and WS small-world model. The power-law distribution means that a very few network nodes were far more connected than other nodes. It indicates the networks can self-organize to a scale-free state, where the highly connected "hub" nodes strongly affect the structure and dynamics of the networks. The average path length of a scale-free network has been found to increase approximately logarithmically with the number of nodes, N. Furthermore, it has also been observed that ℓ of a scale-free network is smaller than a random graph of the same network size and average degree.

In 1999, Barabàsi and Albert (BA) proposed another model to explain the origin of power-law degree distribution. The BA model suggests that the scale-free structure is the consequence of two main ingredients [10]:

1. Growth of nodes: at each time step, a new node is added, with some edges connecting to nodes that already present in the network.
2. Preferential attachment to well connected nodes: the probability of a new node connecting to a node depends on the degree of node i. In other words, a new node is more likely to connect to nodes that are highly connected.

The algorithm of BA scale-free model is as the following: starting with m_0 nodes, and at each time step, add a new node with m edges connecting to pre-existing nodes. The

probability $\prod_i$ that a new node will be connected to node i depends on the degree k_i of node i, in such a way that $\prod_i = k_i / \sum_j k_j$.

2.2.5 FADA in the DBE architecture

The DBE acts as a wide smart Internet-enabled bus that enables services to interact, evolve and integrate. The architecture of the DBE consists of three different environments: the Service Factory (SF), the Execution Environment (ExE) and the Evolutionary Environment (EvE).

Among the three environments, the ExE plays an extremely important role in the DBE. It is where actual services live as they are registered, deployed, searched, retrieved and consumed in this environment. Thus, sometimes it is also referred to as the "run-time of the DBE". It behaves as the gate to access the DBE services. In order to implement the required functions, the ExE comprises a number of core services. The middleware which acts as the connecting glue to them is called the DBE Nervous System. The main role of the nervous system is to store and distribute service proxies. It is implemented based on SUN Microsystems' FADA (Federated Advanced Directory Architecture). FADA is a distributed dynamic directory system based on SUN Microsystems' Jini concept, which maintains services information linking with service registration processes. In this way, changes in the availability of service proxies, can be re ected and service proxies are always synchronized with the process that provides accesses to services.

FADA is based on SUN Microsystems' Jini concept, which is an open architecture that enables developers to create highly adaptive, network-centric services [21]. One successful instance is the application of Jini in the FETISH network [22].

FADA stands for Federated Advanced Directory Architecture. It can be regarded as a virtual lookup server consisting of different FADA nodes that work together to find and access services throughout the network [28]. FADA is a peer-to-peer (P2P) dynamic,

truly distributed system. It holds proxies for services following along the same idea of Jini. But the FADA is designed to work over the Internet, which is the actual DBE working environment.

The structure of the FADA is a graph model, as shown in Fig. 2.3. Once a FADA node is setup, it is ready to be used to store service proxies. When a client sends a request to look for implementation of a service, FADA uses the flooding algorithm to perform the lookup function. An example is provided to describe the flooding algorithm in FADA in detail, as follows.

A client wants to search for the implementation of an interface of a service, for

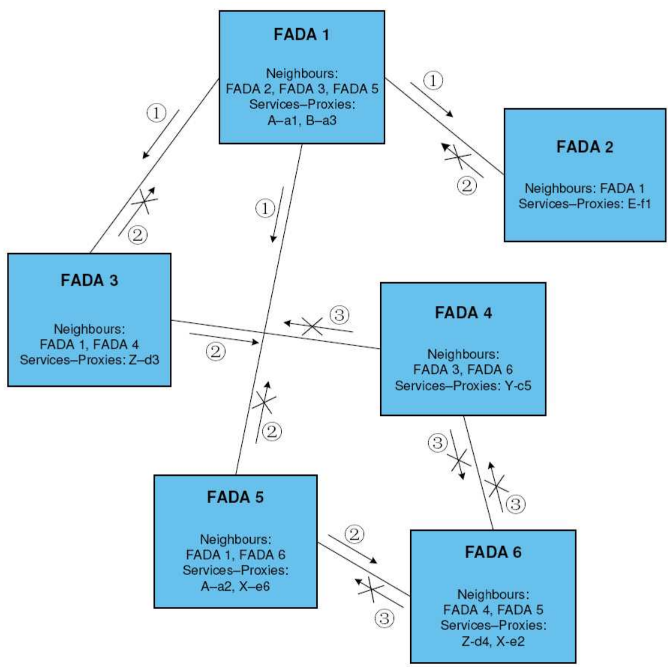


Figure 2.3: An example of detailed architecture of FADA.

instance, Z. It starts from any FADA node, e.g FADA 1, by calling the lookup method in the node. In this case, FADA 1 is regarded as the initiator of the search request. A unique search ID will be created. At first, FADA 1 will look in its own node for the interface and meanwhile, it will propagate the request to its directly connected nodes (FADA 2, 3 and 5). There is no proxy for the interface Z in FADA 1, so FADA 1 will not add any result to

the result set associated with the search ID. In the meantime, FADA node 2, 3 and 5 that receive the search request will perform the internal lookup for interface Z. FADA 3 will find proxy d3 for the interface Z, thus it will send the result directly to the result set in FADA 1. FADA 2 and 5 will not find any proxy for the interface. When these three nodes are executing the internal lookup, they also extend the search to their directly connected nodes respectively. Therefore, FADA 1 will receive the request from FADA node 2, 3 and 5, but it will reject the request as the search ID has already existed in it. FADA 4 and FADA 6 will also receive the search request from FADA 3 and FADA 5 respectively. Then FADA 4 and FADA 6 will do the internal lookup and at the same time, extend the search to their directly connected nodes. There is a proxy for the interface Z stored in FADA 6, so FADA 6 will send the result back to the initiator (FADA 1) directly, instead of via a reverse path. In this way a faster response can be achieved. Consequently, a connection will be created between FADA 6 and the initiator. The search will be extended in the FADA system following this way. A parameter $numHops$ is used to restrict the procedure of the search. It shows how many hops a request can be forwarded, and it is decreased by one on each time of being forwarded. The procedure will be terminated when $numHops$ equals to zero. At the end, the initiator will collect and return the available proxies for the implementation of the requested service to the client. The client can then download the proxies and invoke the methods on the proxies to perform the actual service.

2.2.6 Scope of the work

FADA, which is developed by SUN Microsystems based on its' Jini technology, is implemented as the P2P communication infrastructure to support the business processes between the SMEs and the communications between different parts of the DBE system.

In the DBE architecture with FADA, two problems related to network topology are important to work on. First, what topology would FADA network evolve to as system requirements are changing and distributed mechanisms applied on the network vary. This part of work was done by combining graph-theoretic and system requirements. Second,

how the interactions between the SMEs and FADA communication infrastructure affect the topology of the FADA network. These works are presented in Chapter 3 and 4 respectively.

Chapter 3

Network topology from combining graph-theoretic and system requirements

3.1 Introduction

In the DBE architecture, networks and several superimposed layers of communication appear at several points. The SDC's migrate from the "primordial" soup into agents that will deliver services. At all levels, the software can also move within levels of the DBE. The network distribution infrastructure is superimposed on the network abstraction layer which is aware of the network topology. In the DBE, Federated Autonomous Directory Architecture (FADA) is used for finding services. Based on these assumptions, a two-coupled network model as a simplified version of the DBE is presented. The upper layer is the business network layer where business processes between different SMEs happen. The lower layer is the peer-to-peer (P2P) FADA communication layer that supports communications in the DBE. How the topology of FADA network would change, as system requirements are changing and distributed mechanisms applied on the network vary, is studied. In addition, at the highest level of the DBE, there must be a rationale behind the choice between a random graph, scale free network, etc. Because of the change of

topology, performance of FADA network in the DBE would differ. The study also enables us to further find out how the performance of the P2P network depends on the network topology and dynamics of the network.

3.2 Related work

Recently, Peer-to-peer (P2P) networks have attracted significant attention in both academia and industry due to the widely usage of P2P systems today. Despite of strengths such as the benefit of sharing huge amounts of data, self-organization, and fault tolerance, there are still several challenges for the systems to consider for more widely usage, which are primarily aspects of scalability [23] [35] and efficient search [40] [10] [32] [7] [5] [30]. Moreover, different routing algorithms have also been widely studied to improve performance, e.g in [19], presenting a cost effective broadcast algorithm for fully decentralized P2P networks, implemented on a scale-free topology, based on Barabási and Albert's model in [15]. However, it must be noted that the work mentioned above, the studies of topologies, efficient search techniques and routing algorithms, perform well for the systems in a well-behaved fashion that they are intended for. Their analytical and empirical studies are based on a certain degree of topology. While, in many cases such as in the DBE, the topology of the FADA peer-to-peer network can not be easily derived. It must be highly related to system dynamics and requirements.

Appropriate models of the topological structure of a network are essential for the analysis and development of it. Many researchers have made significant effort in developing topology generators for e.g Internet simulations. Waxman introduced one of the most commonly used methods for graph generation [43], where the probability of an edge between two nodes relates to the Euclidean distance between the two nodes. Due to several serious drawbacks of the Waxman model for generating Internet topology, some modified methods such as [11] have been proposed. In [45], Zegura et al. evaluated different graph generation methods by a comprehensive set of metrics that characterize the graphs produced. However, these methods have not yet been scaled to apply to much

larger size networks. In [18], Paxson and Floyd stated that the difficulty of modelling and simulating the Internet is because the Internet is heterogeneous and changes constantly so that we do not know how to simulate its behaviours. Therefore, in order to provide an appropriate topological model of the FADA P2P network in the DBE, graph-theoretic and system requirements are combined to derive rules for designing a robust and efficient FADA network.

The network of FADA nodes working together is a complex network. There is some research on computer networks that relate to the field of statistical physics and complex networks [9]. Moreover, a lot of studies on complex networks illuminate how specific topologies or connectivity patterns are based on the construction and growth of such networks.

Complex networks are usually regarded as graphs for topological analysis. Similarly, the FADA network can be regarded as an undirected graph where nodes represent FADA nodes, and edges correspond to virtual connections between the nodes. This research relates to some popular topologies of complex networks introduced in chapter 2. These topologies have been applied to real-world research. However, by evaluation, it was found that all these network topologies have their own advantages with respect to different aspects. For example the ER random graph (introduced in chapter 2) is more robust to attacks than a scale-free topology (introduced in chapter 2), while it is less robust to random failures [7]. Thus none of them can be used directly to construct the network. Nevertheless, in this chapter, it will be shown that it is useful to apply their evolution principles as distributed mechanisms to obtain the topology of the DBE and rules for SMEs to communicate with the system. Through a set of numerical analysis, which evolution principles have significant impacts on the system performance can be discovered.

3.3 Methodology

Several topologies for the FADA network have been considered by SUN Microsystems. The first approach was a tree. However, the tree topology as well as several other topologies has some drawbacks. If these topologies are applied to the network, their drawbacks are difficult to overcome. Therefore, the topology of the FADA network was chosen to be un-fixed and unknown in advance. Nevertheless, as the FADA network is applied to the DBE as the communication infrastructure, it is important to investigate how the performance of the P2P FADA communication network is affected by the evolution principles of different topology models. The approach here is to imitate the processes and dynamics of the DBE system in a two coupled networks model. Through simulations on this "run-time" DBE model, the performance of the FADA network under various system configurations with respect to different principles of topologies can be evaluated. This enables possible rules to design a topology for the P2P communication network that is resilient against failure of nodes, adaptable with behaviour of the evolutionary self-organizing system and efficient in obtaining required services to be derived.

In the present work, the flooding algorithm of FADA, is applied to perform the search function in the model in this research. A client sends a request for implementation of a service from any entry point of the FADA and the request is extended to the initiator's neighbours and its neighbours' neighbours and so on. At the end, results are collected and returned to the initiator and then the proxy of the required service can be downloaded by the client. There is a remarkable behaviour during the flooding algorithm that new connections are generated between the initiator and the nodes which receive the request and contain the proxy, if they are not previously connected. This special behaviour is shown to play an important role in the performance of the FADA P2P network in the numerical analysis later.

3.4 Model description

The two coupled networks model as the simplified version of the DBE is presented in Fig. 3.1 below.

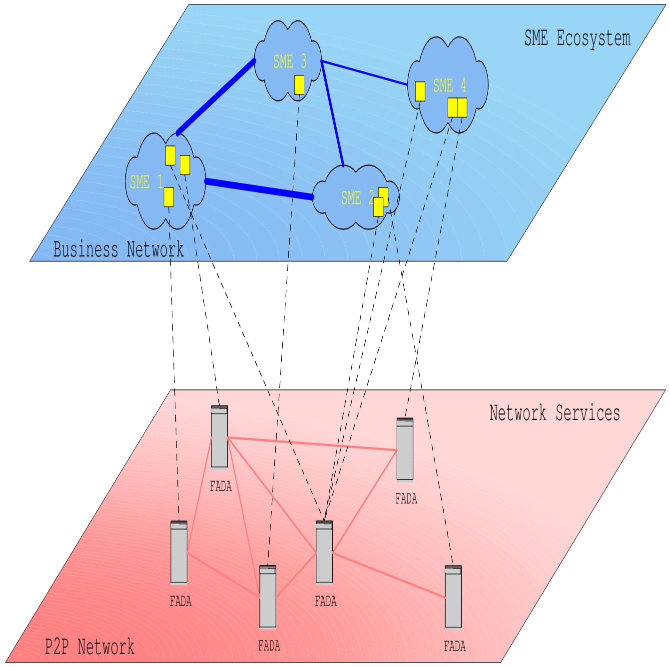


Figure 3.1: Global View of the work

The upper layer is called business network layer, in which SMEs interact with each other for service exchange and migration. The business network layer can be regarded as an undirected graph, in which vertices represent individual SMEs and edges correspond to business connections between them. The lower layer is called P2P communication layer, based on the FADA network. The interactions between these two layers are shown in Fig. 3.2.

The details of communication processes and interactions in Fig. 3.2 are illustrated as follows.

- Once a service provider wants to publish a service, it writes a service proxy which can be invoked to performance an actual service and registers the proxy in at least one FADA node. For instance, in Fig. 3.2, SME B registers one of its proxies in

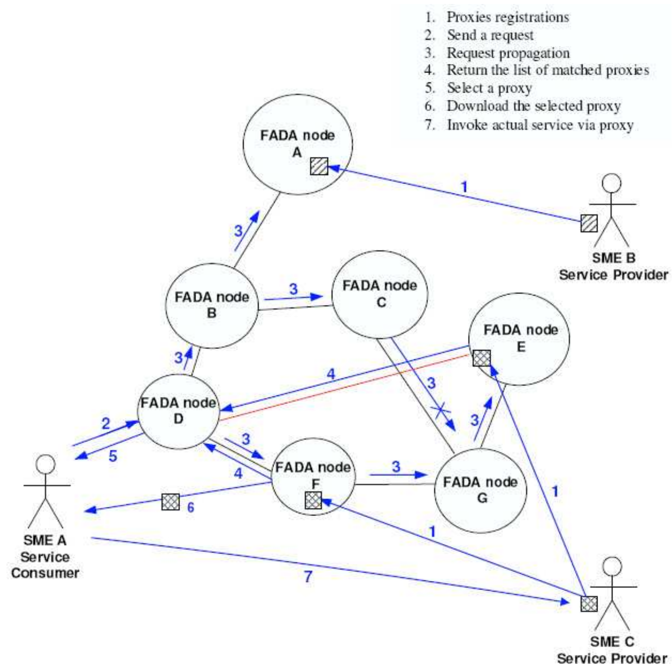


Figure 3.2: Communications in the DBE.

FADA node A, and SME C registers one of its proxies in FADA node E and F.

- When a client SME A, joins the system as a service consumer, it sends a lookup request containing information such as price and functionality of a specific service to a FADA node D, which is called an initiator node of the lookup request.
- The initiator FADA node D makes a lookup in its own data and uses the flooding algorithm to propagate the request in the network. A counter, num_{Hops}, controls how many hops the request can be sent from the initiator FADA D.
- A FADA node E, who finds a matched proxy returns the proxy to the initiator FADA node D via a direct call to the FADA node D, instead of a reverse path so as to allow a fast response. Consequently, a new connection is generated between FADA node E and FADA node D (the new link is represented by a red line in the figure).
- The service consumer SME A selects the best solution among the matched proxies.
- The service consumer SME A downloads the proxy from FADA node F that contains the selected proxy.

- By using the methods on the proxy, the service consumer SME A can communicate with the service provider, which is SME C for the actual service.

SMEs join in and leave the system frequently as in other P2P networks. Some SMEs interact with the system as service consumers and also providers (executing actions 1-7 above). On the other hand, some SMEs who do not have IT infrastructure use the system as service consumers only (being not able to execute action 1).

Based on the model described above, the simulation procedure consists of two main stages which are initialization stage and evolving stage.

In the initialization stage, the model is statically constructed in the following way: At first, the FADA network is setup by using either the ER random graph model or the BA scale-free model. Then assume the number of SMEs in the network is N_{SME}. Among these SMEs, a number of them, n_{SMEReg}, are randomly selected to initially register their services proxies in FADA nodes. Each service proxy is initially registered in n_reg FADA nodes.

In the evolving stage: network topology of a P2P network is always dynamic and changing constantly because nodes continuously enter and exit the system. New nodes can join in and existing nodes can leave at any time. Therefore, the evolution of the two coupled networks in this stage is modelled as below:

- New SMEs join to the network according to the rate n_{joinS}. FADA nodes have been assigned limited memory spaces so that the overload of registries on a single node can be avoided. Each of the new joining SMEs registers its proxies in n_{reg} randomly selected FADA nodes, whose available space is large enough to store the proxy.
- Requests are sent to the system to look for specific services at the rate r_req.
- According to the request information (e.g service type, initiator), FADA nodes work together to look for the request by using the flooding algorithm. The request can

reach other FADA nodes which are n_{Hops} hops away from the initiator.

In our simulation, there are three important aspects, which are setup, registration and new connections for new joining nodes. Following the discussion on the well-known topologies of complex graphs in chapter 2, different evolution principles of random graph and scale-free network are applied. These are to investigate the effects of different distributed mechanisms on the topology of FADA network. The different mechanisms applied in the three aspects are summarized as different cases in the table 3.1 below.

Table 3.1: Different mechanisms used in setup, registration and new connections from new nodes

Case No.	Setup	Registration	New connections from new nodes
1	SF	Preferential	Preferential
2	SF	Preferential	Random
3	SF	Random	Preferential
4	SF	Random	Random
5	RG	Preferential	Preferential
6	RG	Preferential	Random
7	RG	Random	Preferential
8	RG	Random	Random

The objective of the ***setup*** process is to construct a network with $N = 100$ nodes in the initialization stage. *SF* represents the scale-free model. In this case, the network starts with $N = 2$ isolated nodes. At each time step, one new FADA node joins in with $m = 2$ new edges connecting to other FADA nodes, each of which preferentially attaches to a node i with degree k_i, with the probability $k_i/\sum_j k_j$ indicating that nodes with high degree attract more new connections. *RG* represents the ER random graph model that is used, where $N = 100$ initially isolated nodes make connections according to the connectivity probability p. In order to obtain a fair level of comparison, the p has been chosen to be $p = 0.04$ so that after the setup procedure, the total number of edges generated by using *RG* equalling to $p\,N(N-1)/2$, is approximately the same as in *SF*.

Registrations of service interface proxies in a certain amount of FADA nodes by

SMEs happen in both the initialization stage and the evolving stage. ***New connections for new joining nodes*** are only considered when the network enters its evolving stage. *Preferential* mechanism in ***registration*** means proxies are more likely registered in nodes with high degree. On the other hand, *random* means registering proxies in randomly selected nodes. The explanation of the two mechanisms for ***new connections for new joining nodes*** are similar as in the ***setup*** process.

The connectivity of the FADA P2P network can be represented by a $N \times N$ *adjacency matrix* C. The value of an element C_{ij} is either one or zero. The connections between FADA nodes have to be symmetric. Thus, if node i is in node j's neighbor list and node j is also in node i's neighbor list, then $c_{i,j} = c_{j,i} = 1$, or vice versa. This variable is used in the simulations to evaluate the degree distribution and network resilience in the following section.

3.5 Numerical Results

In this section, the simulations results on topological and other related properties, such as degree distribution, average path length and network resilience, are presented. These results can reflect the topology of the network at some level.

As it takes a long time to run the complete simulation and it consumes a lot of computational resources, the results provided here are obtained when the network size evolves to $N = 600$. Several experiments have been run until the network evolves to an even larger size. Based on the results, it has been shown that the results for larger size are predictable from results of a network of relatively smaller size. Moreover, it has been found that when the network size evolves to a large size, e.g. 600, the results are substantially the same if the simulations are re-run.

Simulations are carried out according to the section of model description. The simulation parameters are summarized in table 3.2 below.

Table 3.2: Simulation parameters

Simulation parameters.	values
Number of SME in the initial stage, N_{SME}	100
Number of FADA nodes in the initial stage, N_{FADA}	100
Number of isolated nodes in the scale-free model, n	2
Number of new edges for a new joining node in scale-free model, m	2
Connectivity probability in ER random graph model, p	0.04
Number of SME that register proxies in the initial stage, n_{SMEReg}	20
Number of FADA nodes that one proxy is registered in, n_{reg}	3
Request rate, r_{req}	10
Number of hops in lookup procedure, n_{hops}	5
Rate of SMEs joining the network, r_{joinS}	1
Rate of FADA nodes joining the network, r_{joinF}	1

3.5.1 Degree Distribution

As the most important topological property, the degree distribution of the FADA network is studied by plotting the probability of nodes having degree k, $p(k)$ as a function of k.

Fig. 3.3 and Fig. 3.4 show the results of the FADA network by using scale-free model setup and ER random graph model setup respectively, where (a) to (h) refers to different cases, case 1 to 8 as presented in table 3.1 accordingly.

Power laws have been found in many different fields where networks are applied, comprehensively presented in [9]. Determining whether the FADA network with its evolution behaviour falls into this category is of interest in this research. It is found that the degree distribution of the FADA network roughly follows a straight line on a log-log scale, and most of the nodes have only a few links while only some so-called "hub" nodes have many links. This behaviour is similar to the power-law distribution observed in other people's work. By comparing results in (a) to (h), it can be seen that no matter what mechanisms in the evolution of the network have been used, the degree distributions under different settings are very similar to each other. However, in the FADA network, a "fat tail" in its degree distribution has been found. It is thought that this "fat tail" occurs

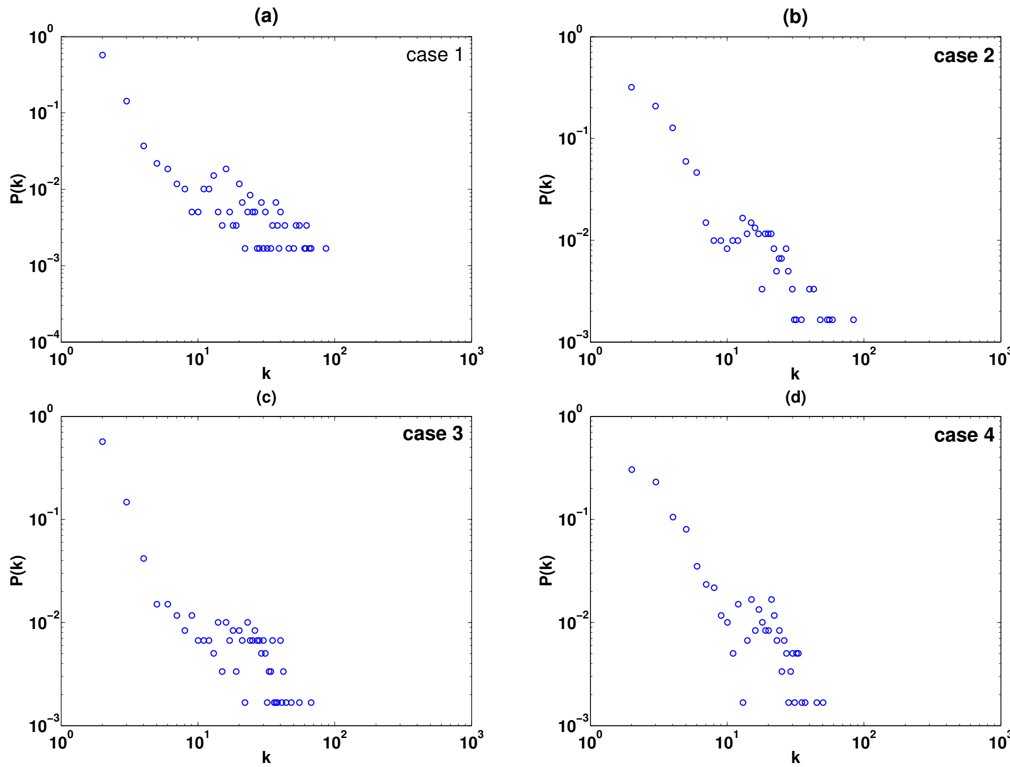


Figure 3.3: Degree distribution $P(k)$ of FADA network under scale-free model setup, with network size of $N = 600$. $P(k)$ is the probability that a randomly selected node having exactly k edges. Referring to table 3.1, case 1: preferential registration and preferential new connections for new nodes; case 2: preferential registration and random new connections for new nodes; case 3: random registration and preferential new connections for new nodes; case 4: random registration and random new connections for new nodes

because of the randomness due to the links generated during the lookup process by using the flooding algorithm.

3.5.2 Average Path Length

In this section, the average path length ℓ of FADA network is studied. Average path length is an important topological property of a network, as discussed in chapter 2. There is no closed formula to compute it yet. But it is widely accepted that this ℓ follows some scaling as a function of a network model's parameters, e.g. size of network N, connection probability p and so on. In the simulation, the Dijkstra algorithm is used to compute the shortest path length ℓ_{ij} between any two nodes: node i and node j. From this the average

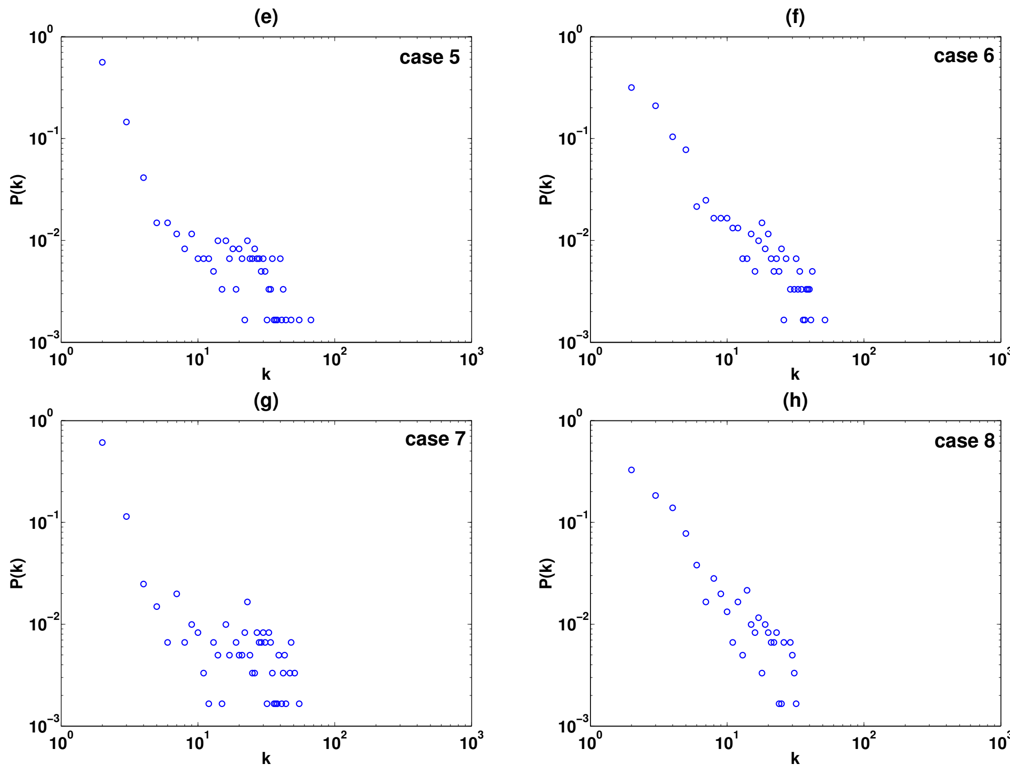


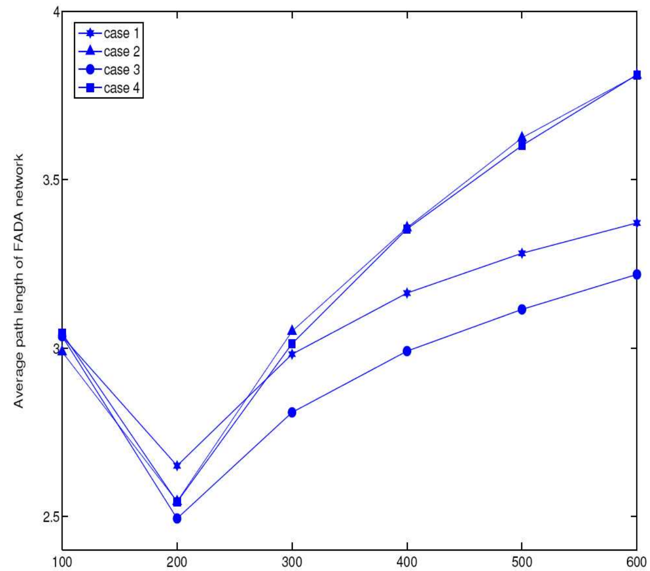
Figure 3.4: Degree distribution $P(k)$ of FADA network under ER random graph model setup, with network size of $N = 600$. Referring to table 3.1, case 5: preferential registration and preferential new connections for new nodes; case 6: preferential registration and random new connections for new nodes; case 7: random registration and preferential new connections for new nodes; case 8: random registration and random new connections for new nodes.

path length for the whole network can be obtained as:

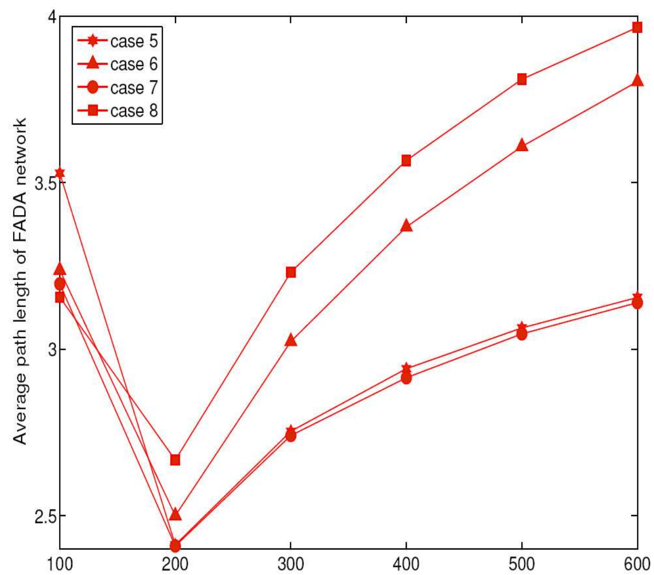
$$\ell = \frac{1}{N^2} \sum_i \sum_j \ell_{ij}, \tag{3.1}$$

where $i, j = 1, 2, \ldots, N$. By observation, it can be found that when the network grows up to 600 nodes, the number of edges in the network becomes very large so that the computation of shortest path is a computationally intensive process. Therefore, the results of average path length of networks of size up to 600 nodes are presented here only. It is believed that, based on the results, the average path length of network when it evolves further can be deduced. The ℓ versus the size of the network N was plotted in Fig. 3.5. Curves in (a) and (b) correspond to the FADA network using SF model setup and ER random graph

model setup respectively.



(a) Scale-free setup: case 1: preferential registration and preferential new connections for new nodes; case 2: preferential registration and random new connections for new nodes; case 3: random registration and preferential new connections for new nodes; case 4: random registration and random new connections for new nodes.



(b) ER random graph setup: case 5: preferential registration and pref- erential new connections for new nodes; case 6: preferential registration and random new connections for new nodes; case 7: random registra- tion and preferential new connections for new nodes; case 8: random registration and random new connections for new nodes.

Figure 3.5: Average path length in the FADA network under different settings referring to cases in table 3.1, with network size of $N = 600$.

Results in Fig. 3.5 show that as the FADA network evolves, the average path length decreases until it evolves to $N = 200$ (corresponding to $t = 100$), ℓ starts to increase monotonically. The critical point $t = 100$ happens for the average path length as the same as for the average number of links studied in the section above. The reason considered is the same as discussed in the section of degree distribution. When $t <= 100$, the average number of connections in the network increases, so the distance between any two nodes is shorter. On the other hand, when $t > 100$, the average number of connections decreases. Consequently, the distance between any two nodes is longer. Moreover, by comparing curves in Fig. 3.5 (a) and (b), it is found that the values of ℓ of networks by using random new connections for new joining FADA nodes in case 2 and case 4 in Fig. 3.5 (a), and case 6 and 8 in Fig. 3.5 (b) are bigger than other cases. Particularly, the ℓ of network under settings of case 8 is the largest among all cases. As in a network, the shorter the average path length, the better performance of finding requested proxies. *Therefore, it can be concluded that the FADA network by using random new connections for new joining nodes displays the worst performance with largest average path length. On the other hand, the FADA network by using preferential mechanism for new connections for new joining nodes, shows better performance due to the comparably smaller average path length.*

3.5.3 Network Resilience

Resilience of networks of different topologies as scale-free network and random graphs may differ. In order to address the resilience of the FADA network by using different distributed mechanisms, so that we can further categorize the topology of the FADA network, the impact of random failures and attacks on the network structure is investigated. The network fragmentation process under failures and attacks is studied by plotting the relative size of the largest cluster S as the function of f, the percentage of removed nodes. The quantity S is calculated by

$$S = \frac{size\ of\ the\ largest\ cluster}{initial\ size\ of\ the\ network}. \tag{3.2}$$

The metric S has been considered as an interesting metric in [25] to show the

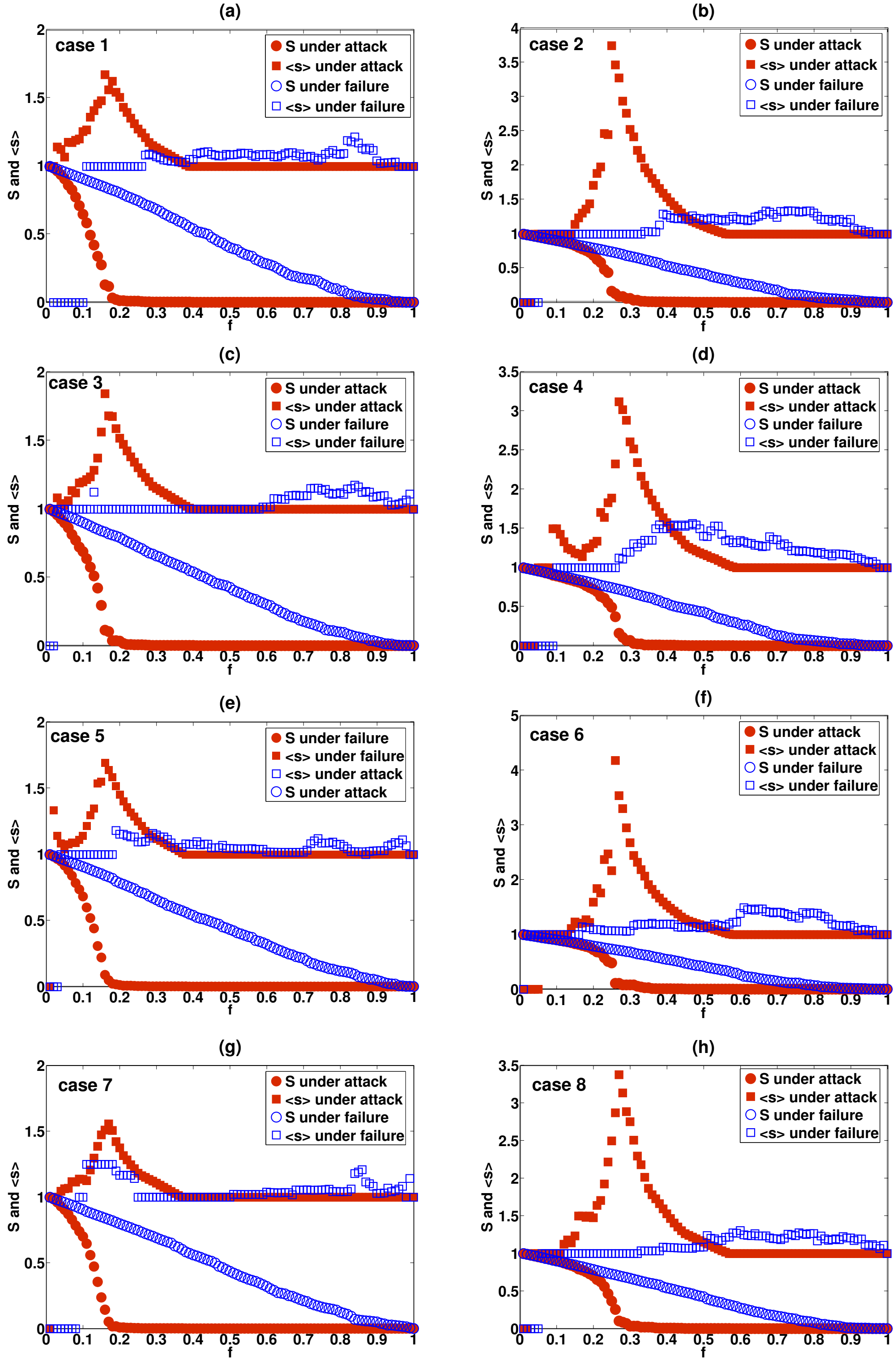


Figure 3.6: Network resilience of FADA under attacks and random failures for different settings referring to case 1 to 8 in table.3.1, with network size of N=600.

maximum number of nodes that can communicate with each other. In addition, we also study the average size of the isolated clusters $\langle s \rangle$ as a function of f. We use the depth-first search algorithm to locate components in different clusters to calculate S and $\langle s \rangle$. The random failures of nodes are simulated by removing randomly selected nodes from the system, while attacks to the system are simulated by removing the most highly connected nodes. Links connecting to these removed nodes are cut off from the network without rewiring. Fig. 3.6 shows the results on the network resilience under different settings as summarized in Table.3.1.

Firstly, we discuss the results on the network under attacks. We find that the relative sizes of the largest cluster S under attacks, presented by filled circles in Fig. 3.6, display the threshold-like behavior for all the settings, which has been observed in [9] [25]. We can see that when we increase the fraction of removed nodes, S decreases steeply. Until $f > f_c$, we have $S \simeq 0$, indicating the network is seriously fragmented when the main cluster breaks into small pieces so that even the largest cluster contains only a few number of nodes. On the other hand, the average size of isolated clusters (presented by filled squares) that fall off from the main cluster, $\langle s \rangle$ reaches its peak value and if we continue to remove nodes, $\langle s \rangle$ starts to decrease, with the result that most of the nodes are from isolated clusters.

By observing the results shown in Fig. 3.6, it is obvious that according to the critical values f_c, we can divide them into two groups as group A: (a), (c), (e) and (g) and group B: (b), (d), (f) and (h). Referring to the table. 3.1, we notice that group A is corresponding to cases that use preferential attachment to make connections for new joining nodes, while group B is corresponding to random attachment. In Fig. 3.6, we can see that f_c of group A is smaller than in group B, where $f_c^A \simeq 0.21$ and $f_c^B \simeq 0.32$. Moreover, we find that $\langle s \rangle_{max}^A < 2$ is always smaller than $\langle s \rangle_{max}^B > 3$. Therefore, we can deduce that by using preferential attachment to generate new links degrades the robustness of the FADA P2P network against attacks because of its smaller threshold f_c of complete fragmentation.

Comparatively, the mechanisms of how to initially setup the FADA model (either by ER random graph model or scale-free model) and of how SMEs register their service proxies in FADA nodes play a less significant role in the resilience of the FADA system against attacks. Despite of this, we can still observe that $\langle s \rangle_{max}$ by using preferential registration is always larger than by using random registration, indicating that using preferential registration also degrades the robustness of the FADA system against attacks.

For the robustness of the system under failure, we find that the system does not display the threshold-like behavior as under attacks. S (presented by circles) decreases slowly and almost linearly with the increasing f, where the values of $\langle s \rangle$ (presented by squares) are between 1 and 1.5 for all the settings, indicating that despite of random failure, the network remains a large cluster with a few isolated clusters consisting of a single node or of two nodes, not a large cluster split into a few major clusters. *Therefore, we may conclude that with all the different settings as in table.3.1, even for the ER random graph setup, random registration and random connections for new nodes, the FADA network is very robust to random failure.* This observation is different from the [9] where the scale-free network is much more robust to random failure than exponential network.

We conclude that due to the significant impacts of links generated during the lookup mechanism, the FADA network has a similar behavior as a scale-free network in the aspect of degree distribution and network resilience under all the settings. Through our simulation, we find that different settings can improve the network resilience to a certain level although they do not change the actual character of the FADA network.

Chapter 4

FADA Topologies based on Evolution-generated communications

4.1 Introduction

4.2 Introduction and background

In the previous chapter, we have studied the topological properties of FADA network by using distributed mechanisms, and the communication implications of service registry decisions. The sophistication in the model was in the interaction of registration process and topology. In this chapter, we incorporate the dynamics that generate the registration of services. These are essentially evolutionary dynamics, generated by the evolutionary environment. Such a combination, of resilient topology, and evolution-generated communications, has never been studied before. The same methods as in chapter 3: simulations inspired by knowledge of graph theory and dynamical systems have been used. These studies enabled us to investigate how the DBE network topology copes with communications generated by the evolutionary environment.

In this chapter, the SME business network is modelled as an information ecosystem and the FADA network is the communication infrastructure that supports business processes in the SME network. How different topologies of the FADA network (which is the Lower Layer, LL) can be obtained by distributed mechanisms that model communications and interactions between SMEs in the upper layer (UL). The dynamic relationship between the topological properties of the FADA communication network and evolution-generated communications due to changes of the topology of SME network is studied.

Computers are now interconnected to a higher degree and in a more complicated way than before, along with increasing demands upon them and improving technologies. Interaction and adaptation of entities in such a complex, dynamic system occur continuously and concurrently. In natural systems, behaviours are unpredictable and undefined, and they always display a high degree of resilience and flexibility. Therefore, it inspires a lot of researchers to apply many aspects of natural systems to different areas. The concept of "information ecosystem" is one of them, which emphasizes the information being exchanged between networked entities rather than the entities themselves [3] [1]. It is titled ecosystem because the information flow in it is an analogy with the flows of materials and energy in a natural ecosystem.

An information ecosystem may represent a wide range of complex systems, as long as there are interactions between information users and providers in the networks. Thus, one way to model the SME network is to regard it as an information ecosystem, in which SMEs manage and communicate with other SMEs. Services and products provided by SMEs are stored in their habitats. Individuals of the SME network can benefit by combining their services with those of others, building new products and services so that they can expand their own business easily.

Interactions between different SMEs are supported and executed by the communication infrastructure, FADA. Many researchers have presented a number of approaches

of information and communication infrastructures to support business processes between different business enterprises, such as CORBA [4], Java RMI [36] and DCOM [2], providing basic information infrastructures. Another enhanced infrastructure which consists of several replicable e-business servers for collaborative e-business applications is presented in [34]. In [12], several adaptive methods and rules for routing messages are applied to construct a self-adaptive infrastructure for business process.

All the infrastructures and methods discussed above emphasize primarily how to support and improve the integration of cooperative business processes among multiple enterprizes. However, in the field of business process reengineering (BPR), most of the researchers suggest that information infrastructure plays an important role in the reengineering process. Despite business process and information infrastructure interact in practice, efforts on BPR and information infrastructure are rarely conducted together. But obviously, the study of business processes can not be isolated from information infrastructure. It is important to explore and understand the dynamic relationships between the two and explore the effects of changes of one of these on the other. Some practitioners have investigated this thinking [12] [13] [33]. However, further analysis on this relationship in more detail is needed.

In this chapter, a two coupled networks model which is more complex than the model presented in the previous chapter is proposed. The discrete-event simulation tool is used to model the dynamics and behaviours of a simplified version of the real-world DBE system. In this way, a comprehensive analysis on how different topologies of the FADA network are affected by topologies of the SME network and the dynamics related to the business process in the SME business network can be achieved.

4.3 Model description

In the previous chapter, a two coupled networks model was presented. In that model, it was considered that all FADA nodes were available to all SMEs. Thus every SME was

able to register its services in whichever FADA node. This was for the purpose of making a simple assumption so that the effects of different network topology mechanisms and service registry decisions on the performance of the FADA network could be studied. However, in the real world, capacities and bandwidths of links between nodes located in different geographical regions are various. Moreover, some nodes may be unavailable to some users. These considerations do not affect the findings that were obtained in the previous chapter. Nevertheless, this needs to be taken into consideration here. The reason is that the focus of the work in this chapter is about the effects of communications and interactions between SMEs on the topology of the FADA network. Therefore, it is important to model how the two networks interact appropriately. In this model, communications generated by evolutions of populations of services in the evolutionary environment and migrations of copies of services between habitats of SMEs, are considered. Moreover, the registration of services is also driven by the communications to implement the interactions between the UL and LL. Therefore, how SMEs communicate with each other and with FADA nodes is an important factor. A modified model is illustrated in the Figure 4.1.

Business interactions happen more frequently between SMEs within the same geographical region, therefore SMEs are modeled to be virtually grouped according to their geographical locations. As FADA nodes in the communication infrastructure support the business processes of SMEs, they are also virtually grouped depending on the locations of the SMEs that they are serving. SMEs in a region can only communicate with FADA nodes which are in the same region as them.

In the initialization stage, the mechanism used to setup the FADA network is by using the ER random graph model with a fixed connectivity probability p_{FADA}. As the communications between SMEs are dependent on the topology of the network of SMEs, the network of SMEs are modeled by using the three most popular topologies in the field of complex network introduced in chapter 2, which are ER random graph model, BA scale-free model and small-world network model. SMEs initially register their services in

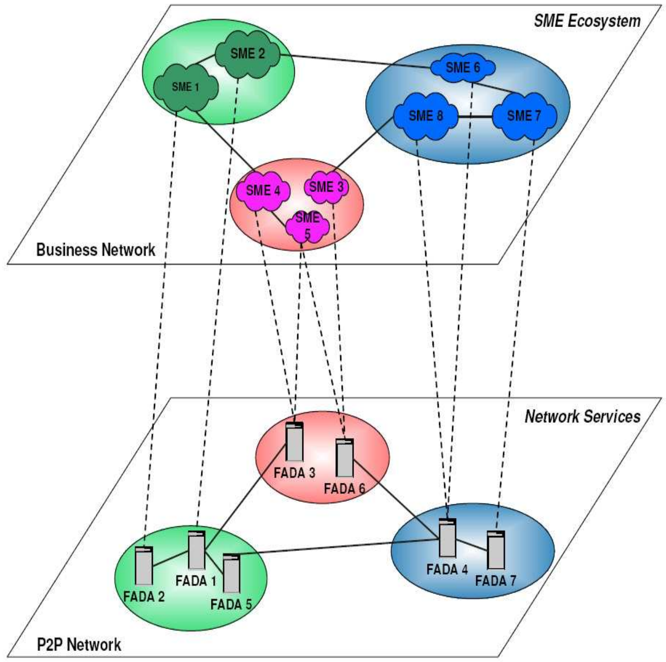


Figure 4.1: Architecture of the two coupled network of the DBE.

the FADA nodes which are available to them.

In the evolving stage, at each time step, each FADA node checks leases of proxies that are registered in it in order to know how long a service proxy can be kept in the FADA node. If the lease of a proxy is to expire, the node has to make a decision whether to renew it or not, based on certain rules.

As time evolves, SMEs join and leave the system frequently. The SMEs are modeled joining the system at a rate. When a new SME joins in the system, a FADA node is generated to serve it accordingly. In order to maintain the topology of the FADA network as a ER random graph model with the fixed connectivity, a number of links links are made connecting the new joining FADA node and some randomly selected FADA nodes. The new SME registers its services proxies in randomly selected FADA nodes which are available to them.

In the simulation, requests are generated at a rate per time step. When a request

is sent to, for example SME i, an evolving population in the service habitat of the SME i is activated. Service habitats of SMEs are the places where copies of pointers to services provided by their own SMEs and also other SMEs are stored. According to the requirements of the request, fitness values of different services are computed by using a fitness function. Through a set of simple genetic algorithm, an optimal solution is created. For the purpose of exchanging and sharing good services between SMEs, copies of pointers to services in the optimal solution generated from the service habitat of SME i are migrated to the service habitats of the SMEs which are neighbors of the SME i. For those habitats of SMEs that the copies of pointers have been migrated to, the migrated copies result in an evolution of populations of services in the habitats. The strategy of the evolution is that popular services will reproduce and less-used services will die out. Accordingly, the registrations of proxies of these services in the FADA network will be changed. The optimal solution, which contains the IDs of the selected services, is then sent to a randomly selected FADA node available to the SME i. The flooding algorithm of FADA is then performed in the FADA network to look for appropriate proxies for the selected services. With the proxies, SME i can invoke actual services from different SMEs as providers by using the Java RMI.

In the SMEs business network of the two coupled networks model, SMEs exchange services with each other in order to achieve better profit. Inspired by evolutions of species in natural ecosystems, where fit species reproduce and weak species die out, populations of different services in SMEs can also evolve according to the fitness values of different services in this way. In order to define the rules for the evolution of populations of services, replicator dynamics in evolutionary game theory has been applied to the model.

In the SMEs business network layer, each SME owns a service habitat, in which services of different business types exist. For instance, for a SME which is a travel agency, its habitat may maintain services such as airline, car rental, hotel, restaurant, cleaning services and so on. The providers of these services can be the SME itself or by other SMEs. When a request is sent into a SME, some services in the SME are selected to

construct a service chain. The selection of services is based on their fitness values, which are depending on the differences between services and the request and also their usage histories. The usage history of a services means the number of times that has been used in the past. Therefore, if a SME can be guided by some rules so that it knows how to keep and reproduce services of high fitness values and get rid of services of low fitness values, solutions that the SME can offer may be better, benefiting not only the individual SME but also the whole DBE system.

In the model, the optimization of solutions to requests are also taken into consideration. When a request is sent to a SME, the SME creates an evolving population, in which corresponding services are aggregated to construct a number of service chains to the request. A simple optimization process by using crossover genetic algorithm is applied to the service chains generated in the evolving population in order to provide the optimal solution to the request. The procedure of answering a request in the DBE is illustrated in Fig. 4.2.

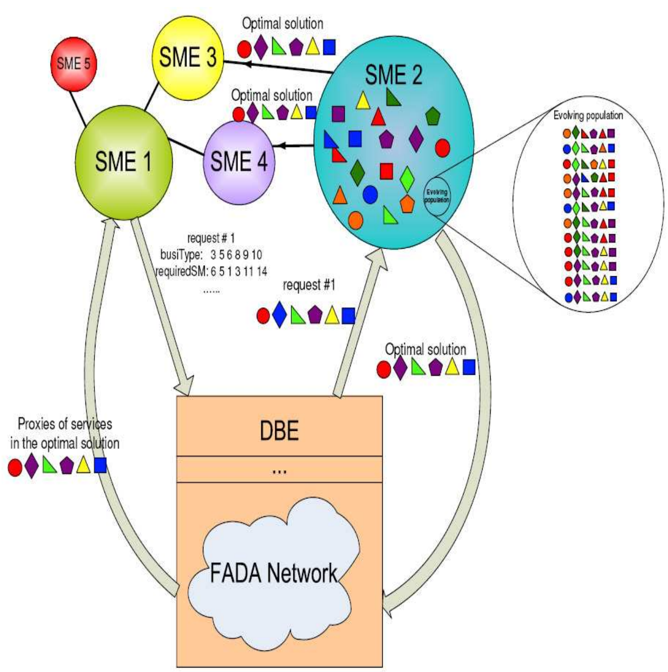


Figure 4.2: The average degrees of FADA nodes in the FADA networks which are coupled with the SME business network of different topologies.

The example of a request in the previous section is used here to demonstrate how solutions to the request are generated. For instance, SME 1 sends a request 1 to the DBE system. The request is in the form of a service chain, and it contains the information of the types and the service manifest descriptions of the required services, which are represented by “busiTypes: 3 5 6 8 9 10 ” and “requiredSM: 6 5 1 3 11 14” respectively in Fig. xxx. The DBE system sends the request to a suitable SME, e.g. SME 2. In the figure, each shape represents the business type of a specific service, and each colour represents the service manifest descriptions of a specific service. SME 2 receives the request 1 from the DBE system, and it generates an evolving population corresponding to the request 1. In the evolving population, it first collects all the services whose business types are the same as the required services in the request, but these services are of different service manifest description. Then the SME 2 generates a number of possible solutions by aggregating these services into service chains. Due to the memory limitation of SMEs, the number of solution chains generated in the evolving population in SME 2 is bounded by S = 1000. SME 2 computes the fitness value of each solution chain. The solution with the highest fitness value is selected as the optimal solution to the request 1, and is sent to the DBE. The FADA network then provides the proxies of the services in the optimal solution to SME 1, where the request is originally sent from. The optimal solution is generated from the evolving population regarding to the fitness value of the solution, through a one point crossover genetic algorithm.

4.4 Empirical results

4.4.1 Topology of the FADA network affected by the connectivity of SME network

Based on the model description, discrete-event simulations have been carried out. Primary simulation parameters are summarized in table 4.1.

Communication flows generated due to the migration of copies of pointers to good

Table 4.1: Parameters and the values of them in the simulations

Simulation parameters.	values
connectivity probability of SME network by using ER model, p_{SME},	0.2,0.4,0.6,0.8
connectivity probability of FADA network by using ER model, p_{FADA},	0.2
initial space assigned to each ADA node, s_{ini}	64,000
maximum size of one proxy, s_{proxy}	200
maximum lease time for a proxy, $lease_{max}$	100
maximum number of hops in lookup procedure, $hops_{max}$	1,2,5
the number of FADA nodes for registrations of one proxy, n_{reg}	1,3,5
rate of sending requests, r_{req}	1,2,3,4,5,6

services and the changes of registrations of services proxies are depending on the topology of the SME business network. Therefore, it is interesting to study how the topology of the SME business network (UL) affects the topology of the FADA network (LL).

In the simulation, ER random graph model is applied to both the SME network and the FADA network. Due to the objectivity of the study of this section, the connectivity probability of the FADA network is fixed as $p_{FADA} = 0.2$, while the connectivity probability of the SME network p_{SME} is of various values as shown in table 4.1. The idea of replicator dynamics has been applied to the evolution of populations of services in the model.

As described in the section of model description, the migration and the evolution of populations of services happen at all time, services that are kept in any SME habitat are changing accordingly. The evolution of populations of services in the UL results in changes of registrations of services proxies in the LL. How different values of p_{SME} affect the topology of the FADA network on the topology of the FADA network are quantified by topological properties including average degree, average path length and network diameter was studied. Firstly, the average degree of FADA nodes as time evolves was plotted when p_{SME} are set as various values in Fig. 4.3. In the simulation, at every 10 time steps, one new FADA node is generated and new connections originated from it are made according to the ER random graph model. Therefore, theoretically, the average degree of FADA network should increase along with the network size grows as presented

in the random graph theory [1]: $\langle k \rangle = p \times N$. Here p is $p_{FADA} = 0.2$. However, it is shown in Fig. 4.3 that the average degree of FADA nodes increases much faster than k_{RG} when $p = p_{FADA} = 0.2$. This indicates the significant impact of new connections generated during FADA network's lookup procedure. Moreover, it can be noticed that as p_{SME} increases, the increase of the average degree of FADA nodes is faster. The explanation of this is as the following: larger p_{SME} results in more communication flows between SMEs and migrations of services in SMEs' habitats from a wider range of SMEs. As described in the section of model description, SMEs can only register their services in several local FADA nodes that are available to them. When an optimal solution is sent to the FADA network to look for actual proxies, if the solution consists of services originally exist in SMEs of a wider range, initiator FADA node which initiates the lookup request may find proxies in registered FADA nodes of a wider range accordingly. Consequently, more new connections may be made to the initiator when nodes return results directly to it. Therefore, the larger the p_{SME}, the higher the average degree of FADA nodes is. As the average degree of FADA network with larger p_{SME} is higher, it can be conjectured that the average path length is shorter in this case. It is confirmed by the simulation result shown in Fig. 4.4.

Furthermore, in Fig. 4.4, it can also be seen that, the average degree of FADA nodes increases so fast that the average path length drops steeply to some very small values after a short time, which are between 1.2 and 1.6. Another important topological property of a network, the network diameter, which is defined as the maximal distance between any pair of nodes has also been studied. It is an important character of a network because it shows the worst case for two nodes to communicate with each other. The network diameter of the FADA network d_{FADA} against time was plotted in Fig. 4.5. For various values of p_{SME}, it is shown that d_{FADA} always equals to 2 approximately, which implies that the FADA network is very efficient for communication in all circumstances.

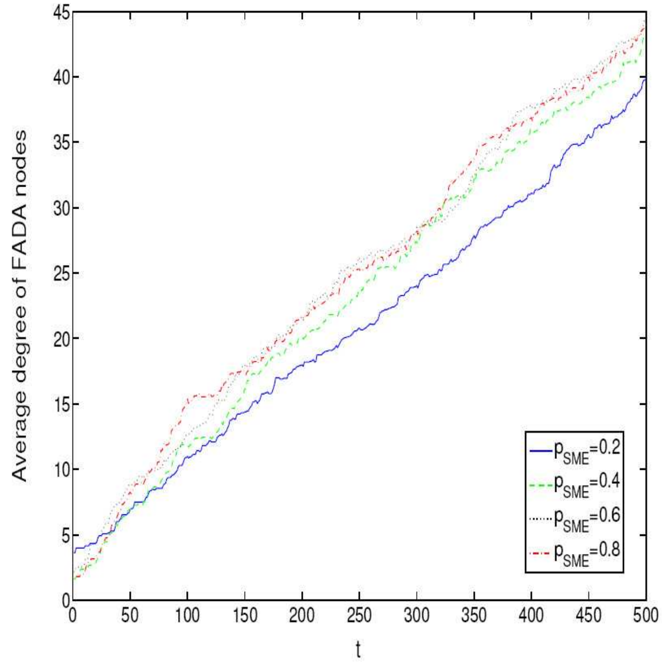


Figure 4.3: The average degrees of FADA nodes in the FADA networks which are coupled with the SME business network of different connectivity probabilities.

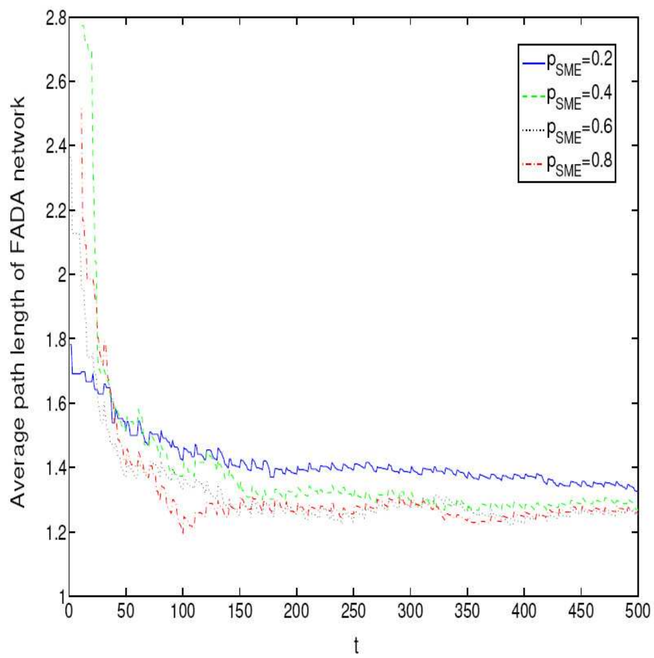


Figure 4.4: The average path lengths of FADA nodes in the FADA networks which are coupled with the SME business network of different connectivity probabilities.

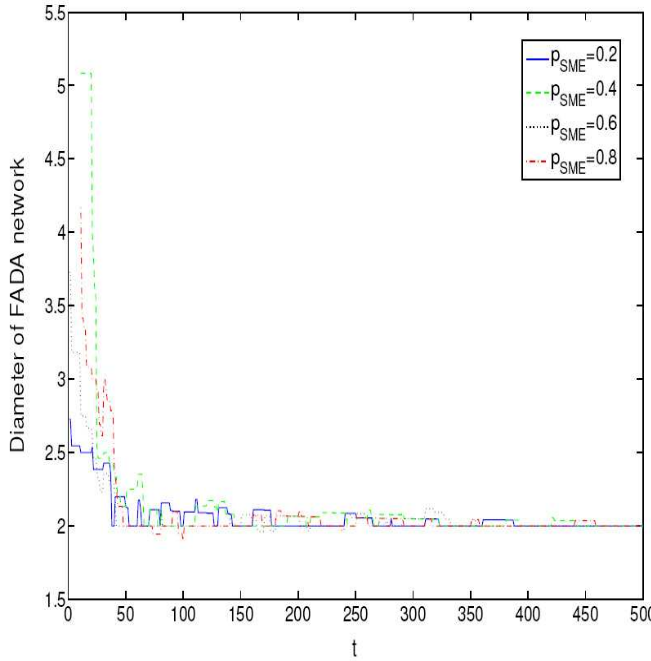


Figure 4.5: The network diameters of FADA nodes in the FADA networks which are coupled with the SME business network of different connectivity probabilities.

4.4.2 Topology of the FADA network affected by models applied on the SME network

In the previous section, the effects of using various connectivity probabilities p_{SME} of the ER random graph model for the topology of the SME network on the topology of the FADA network have been discussed. Changing the p_{SME} reflects the impact of communication flows between SMEs on the system directly. Besides that, another interesting and important thing investigated is how different topologies of the SME network affect the FADA topology. Three dominating topologies in the territory of complex networks which are random graph, scale-free network and small-world network have been applied to the SME network. An analysis of the effects of using these three different topologies is provided in this section.

In order to make a fair comparison between the three topologies, the networks are modelled to have the same average degree and the number of edges. The methodology is to make a fixed number of new connections for a new node when it joins in the

network for all the three topologies. For (1) random graph network: new connections link to randomly chosen existing nodes, (2) scale-free network: new connections are made according to the "preferential attachment" algorithm, which means the placements of new connections depend on the degrees of existing nodes. Introduction to the random graph model, scale-free network model, and small-world network model have been provided in chapter 2. How to use evolution principles of random graph and scale-free network to construct a model has been described in the previous chapter. In order to construct a small-world network, the mechanism used in this simulation was inspired by a simple mechanism proposed by Ozik, Hunt and Ott for the evolution of small-world networks in [124], which is, when a new node joins in the network, it connects only to existing nodes that are geographically close to it. It is a very realistic model for the SME business network, as SMEs locate in different geographical regions and the network size of the SME network is supposed to grow gradually. Therefore, this simple mechanism is applied to construct the small-world topology for the SME network. The procedure is as the following: Firstly, in the initial stage in the simulation, it is aimed to build a SME network of ten nodes. It starts with three fully connected nodes and a new node is added in the network which is placed in a randomly chosen geographical region. A new node makes two connections to its two nearest neighbors according to their locations. It continues until the network size grows to ten nodes. Then in the evolving stage in the simulation, the network grows according to the following steps: (1) a new node joins in the network at rate r_{join} and it is placed in a randomly chosen location (2) the new node connects to eight nearest neighbors of it. The result of applying this mechanism is a network with a small average path length of 2.01 and a large clustering coefficient of 0.59, which satisfied the requirement of being a small-world network.

The impacts of applying different topology models to the SME network on the topological properties of the FADA network were studied, including the average degree, average path length and network diameter. In Fig. 4.6, it is noticeable that when the SME network is constructed by using the small-world network model, the average degree of FADA nodes is smaller than by using the random graph model and the scale-free network

model. Consequently, the average path length of the FADA network is larger when the SME network is of small-world network topology, as shown in Fig. 4.7. Moreover, the result in Fig. 4.8 shows the network diameter of the FADA network drops to 2 after a short while for all the three cases. This implies that the FADA network can provide good communication efficiency no matter which topology model is used for the SME network.

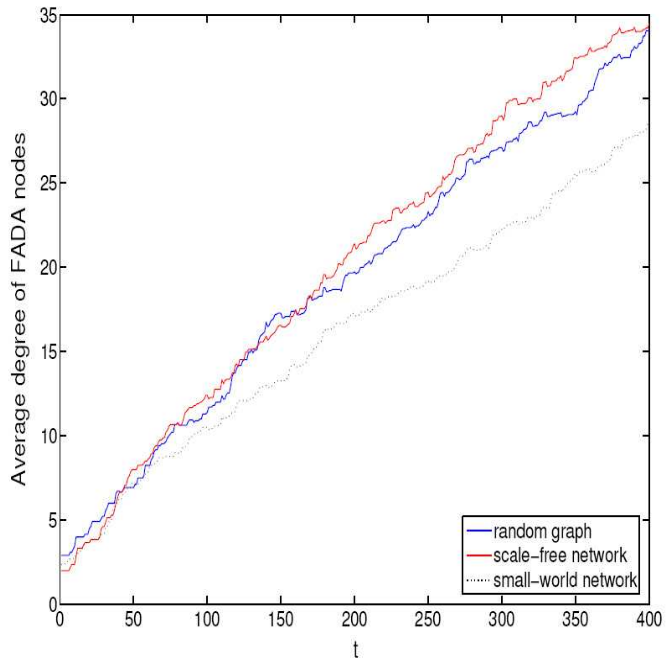


Figure 4.6: The average degrees of FADA nodes in the FADA networks which are coupled with the SME business network of different topologies.

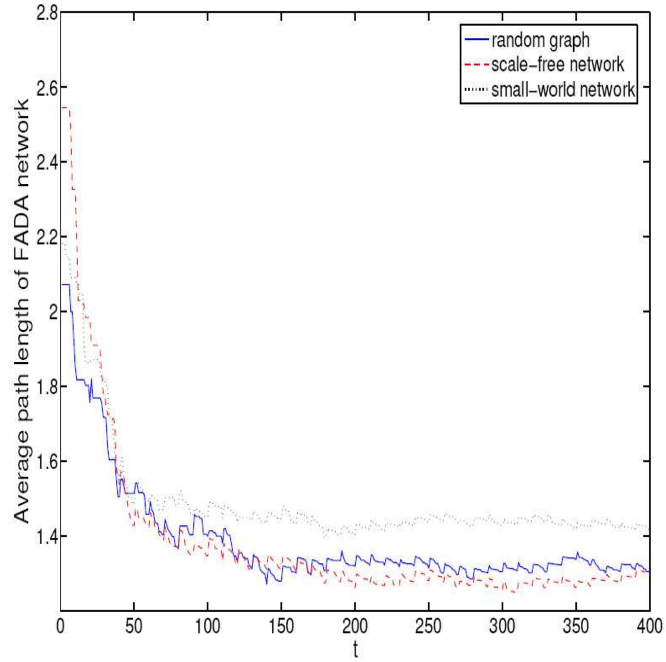


Figure 4.7: The average path lengths of FADA nodes in the FADA networks which are coupled with the SME business network of different topologies.

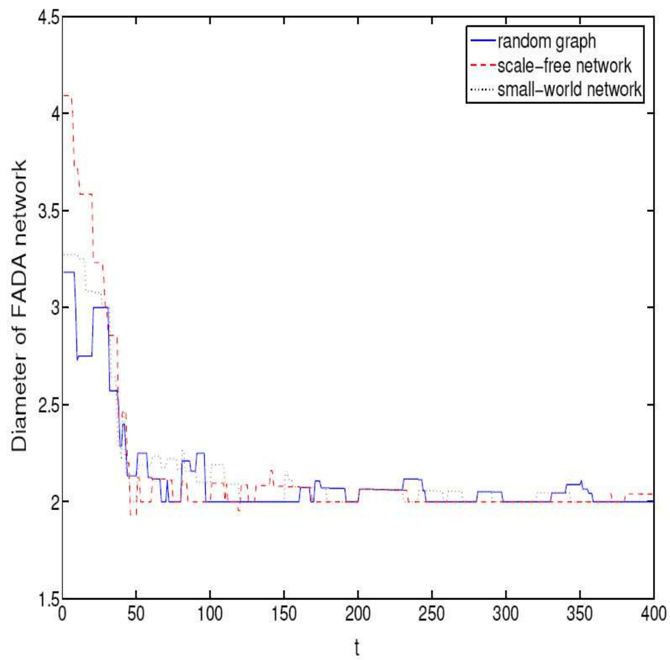


Figure 4.8: The network diameters of FADA nodes in the FADA networks which are coupled with the SME business network of different topologies.

Chapter 5

Summary

The DBE is an all-inclusive environment, where SMEs can leverage the best possible solutions from the Internet technology to expand business of SMEs via latest technology, information and business tools. In order to fulfill the promises that the DBE is to offer, many aspects have been taken into consideration when the DBE approach is proposed to be applied to real world SMEs. These aspects include security, trust, sharing, and competitiveness and so on.

For instance, one of the reasons that the DBE approach has enormous value is that it allows the sharing of results and integration technology in an open-source environment. But it is interesting to know whether the DBE itself can promote co-operation. This relates to the emergence of symbiosis in an ecosystem. Through symbiosis and cooperation, business can perform transactions that they would be unable to perform by themselves. In the DBE project, the effects that the assembling services can have on business relationships are studied in [129]. These studies can help to predict the impact that participation in the DBE will have on regional economics. In addition, how sharing of and competition for services affects the efficiency of the market are also investigated.

In this report, research work on the DBE are mainly concerned with how the topologies of the FADA network are affected by distributed mechanisms and evolutionary principles

inspired by topological models, interactions and communications between the SMEs and different topologies of the SME network.

A two coupled networks model has been constructed and presented as a simplified version of the DBE system in chapter 3. The upper layer is the business network layer where business processes between different SMEs happen. The lower layer is the Peer-to-peer (P2P) Federated Advanced Directory Architecture (FADA) communication layer that supports communications in the DBE, which was developed by SUN Microsystems. Inspired by studies on different topologies in the territory of complex networks, the effects of different evolution principles of random graph and scale-free networks on the topology of the FADA network has been examined through numerous simulations. Results show that network properties such as degree distribution, resilience are strongly affected by different mechanisms applied. The findings can be summarized in the following points.

- *Degree distribution*: it can be concluded that whatever the initial setup of the network, either by random graph model or scale-free model, there is little impact on degree distribution. It was observed the power-law distribution as in scale-free network in all the simulations, but with a "fat tail".

- *Average number of links*: it was found that the average number of links of nodes increases steeply as time elapses until at approximately $t = 100$, E starts to decrease exponentially.

- *Average path length*: it was observed that the average path length ℓ decreases until the network evolves to $N = 200$ (corresponding to $t = 100$), ℓ start to increase monotonically. In addition, the FADA network by using random new connections for new joining FADA nodes displays the worst performance with largest average path length. On the other hand, the FADA network by using preferential mechanism for new connections for new joining nodes, shows better performance due to the comparably smaller average path length.

- *Network resilience*: it was found that the network is very robust against random

failures of nodes. Moreover, the network resilience against attacks can be improved if preferential attachment for new connections from new joining nodes is used. The mechanisms of how to initially setup the FADA model (either by ER random graph model or scale-free model) and of how SMEs register their service proxies in FADA nodes play a less significant role in the resilience of the FADA system against attacks.

In chapter 4, a modified two coupled networks model of the DBE was presented. The SME business network is regarded as an information ecosystem, where behaviors of business processes such as exchanges of services between SMEs and optimizations of service solutions are modelled. This network is essentially an evolutionary environment that generates evolutionary dynamics. As business processes between SMEs happen all the time and the communications are supported by the FADA network, the SME network and the P2P FADA network interact with each other and evolve over time.

A comprehensive study on how communications generated by the SME network impact on the topology of the FADA network is provided. The communications of the SME are highly related to the graph connectivity of the SME network and different topologies of the SME network. The main findings can be summarized as follows:

- The graph connectivity of the SME network matters: the SME network is constructed and evolved based on the ER random graph model with various connectivity probability p_{SME}. With larger p_{SME}, the fitness values of solutions that the DBE system can offer are higher. Furthermore, it was found that the scaling law between the p_{SME} and the usage records of services in SMEs is linear. The results also showed the effects of p_{SME} on the topology of the FADA network. With larger p_{SME}, the average degree of FADA network is higher, which results in smaller average path length. However, different values of p_{SME} have insignificant effects on the diameter of the FADA network.
- The topologies of the SME network affect the topology of the FADA network: three most popular models of topologies in the filed of complex networks are applied to the SME network, which are the random graph, the scale-free network and the small-

world network. It has been shown that by using small-world model for the SME network, the average degree of the FADA nodes is smaller and the average path length of the FADA network is larger than by using the other two topology models for the SME network. Nevertheless, he effect on the diameter of the FADA network is insignificant.

Bibliography

[1] http://www.freenetproject.org/, last accessed: october 2005 edition.

[2] http://www.microsoft.com/com/tech/dcom.asp.

[3] *DIET Web site*. http://www.dfki.uni-kl.de/IVSWEB/DIET/public/index.html.

[4] *Object Management Group. The Common Object Request Broker: Architecture and Specification*, 3.0 edition edition, July 2002.

[5] Lada A. Adamic, Rajan M. Lukose, Amit R. Puniyani, and Bernardo A. Huberman. Search in power-law networks. *Physical Review E*, 64(046135), September 2001.

[6] R. Albert, H. Jeong, and A.-L. Barabàsi. Internet: Diameter of the world-wide web. *Nature (London)*, 401:130 – 131, September 1999.

[7] Reka Albert and Albert-Làszlò Barabàsi. Topology of evolving networks: Local events and universality. *Phys. Rev. Lett.*, 85(24), 2000.

[8] Reka Albert and Albert-Làszlò Barabàsi. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47, January 2002.

[9] Reka Albert, Hawoong Jeong, and Albert-Làszlò Barabàsi. Error and attack tolerance of complex networks. *Nature*, (406):378–382, 2000.

[10] Albert-Làszlò Barabàsi and Reka Albert. Emergence of scaling in random networks. *Science*, 286:509, October 1999.

[11] Kenneth L. Calvert, Matthew B. Doar, and Ellen W. Zegura. Modeling internet topology. *IEEE Communications magazine*, pages 160–163, June 1997.

[12] Yves Caseau. Self-adaptive middleware: supporting business process priorities and service level agreements. *Advanced engineering informatics*, 19:199–211, May 2005.

[13] T. H. Davenport. *Process innovation: reengineering work through information technology.* MA: Havord business school press, Boston, 1993.

[14] S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks. *Adv. Phys.*, 51:1079–1187, 2002.

[15] S. N. Dorogovtsev and J. F. F. Mendes. *Evolution of Networks - from biological nets to the Internet and WWW.* Number 5. Oxford University Press, New York, USA, 2003.

[16] Paul Erdös and Alfrad Rènyi. On the evolution of random graphs. *Publ. Math. (Debrecen)*, 6:290, 1959.

[17] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *Comput. Commun. Rev.*, 29(4):251–262, Ocotober 1999.

[18] Sally Floyd and Vern Paxson. Difficulteis in simulating the internet. *IEEE/ACM transactions on networking*, 9(4):392–403, August 2001.

[19] M. K. Hill. *Understanding Environmental Pollution.* Cambridge University Press, Cambridge, 1997.

[20] B. Huberman and L. Adamic. Internet: Growth dynamics of the world-wide web. *Nature*, 401:131, September 1999.

[21] Marco Iansiti and Roy Levien. Strategy as ecology. *Havard Business Review*, pages 68–78, March 2004.

[22] J.F.Moore. *The Death of Competition.* HarperBusiness, 1996.

[23] A. S. Klovdahl. Social networks and the spread of infectious diseases: The aids example. *Social Science & Medicine*, 21(11):1203–1296, 1985.

[24] Chunguang Li, Hongbing Xu, Xiaofeng Liao, and Juebang Yu. Synchronization in small-world oscillator networks with coupling delays. *Physica A*, 335(3-4):359–364, April 2004.

[25] Damien Magoni. Tearing down the internet. *IEEE Journal of selected areas in communications*, 21(6):949–960, August 2003.

[26] Sun Microsystems. *DBE deliverable 11.2, DBE nervous system.*

[27] Stanley Milgram. The small world problem. *Psychology Today*, 61, 1967.

[28] Francesco Nachira and et.al. Towards a network of digital business ecosystems fostering the local development. *Discussion paper, Digital Busienss Ecosystem Project*, September 2002.

[29] Theoden I. Netoff, Robert Clewley, Scott Arno, Tara Keck, and John A. White. Epilepsy in small-world networks. *The Journal of Neuroscience*, 24(37):8075–8083, September 2004.

[30] M. E. J. Newman, Stephanie Forrest, and Justin Balthrop. Email networks and the spread of computer viruses. *Physical Review E*, 66:035101, 2002.

[31] S. Redner. How popular is your paper? an empirical study of the citation distribution. *Eur. Phys. J. B*, 4(2):131–134, 1998.

[32] F. Saffre and R. Ghanea-Hercock. Simple laws for complex networks. *BT Technology Journal*, 21(2), April 2003.

[33] Cheri Speier, Michael Harvey, and Jonathan Palmer. Implementing intra-organizational learning: an infrastructure to support business process reengineering. *Knowledge and process management*, 5(2):76–86, 1998.

[34] Stanley Y. W. Su, Herman Lam, and Minsoo Lee. An information infrastructure and e-services for supporting internet-based scalable e-business enterprises. *5th International Enterprise Distributed Object Computing Conference*, September 2001.

[35] Ke Sun and Zhen-Xiang Han. Analysis and comparison on several kinds of models of cascading failure in power system. *Transmission and Distribution Conference and Exhibition: Asia and Pacific, 2005 IEEE/PES*, pages 1–7, August 2005.

[36] Sun Microsystems, http://java.sun.com.

[37] UBHAM. *DBE deliverable 11.1: Report on the flow of software components in a static network*.

[38] UBHAM. *DBE deliverable 11.2: Report on clustering in networks*.

[39] UBHAM. *DBE deliverable 11.3: Report on hierarchical networks*.

[40] Xiao Fan Wang and Guanrong Chen. Complex networks: small-world, scale-free and beyond. *IEEE Circuits and Systems Magazine*, 3(1):6–20, 2003.

[41] Duncan J. Watts. *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press, 1999.

[42] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small world' networks. *Nature (London)*, 393:440–442, 1998.

[43] B. M. Waxman. Routing of multipoint connections. *IEEE Journal on selected areas in communications*, 6:1617–1622, December 1988.

[44] Damián H. Zanette. Dynamics of rumor propagation on small-world networks. *Physical Review E*, 65(041908), March 2002.

[45] Ellen W. Zegura, Kenneth L. Calvert, and Michael J. Donahoo. A quantitative comparison of graph-based models for internet topology. *IEEE/ACM transactions on networking*, 5(6):770–783, December 1997.

[46] Nouredine Zekri and Jean Pierre Clerc. Statistical and dynamical study of desease propagation in a small-world. *Physical Review E*, 64(056115), November 2001.