Congestion control and network management in Future Internet

MÁRTON CSERNAI, ANDRÁS GULYÁS, ZALÁN HESZBERGER, SÁNDOR MOLNÁR, BALÁZS SONKOLY

Budapest University of Technology and Economics, Department of Telecommunications and Media Informatics

{csernai, gulyas, heszi, molnar, sonkoly}@tmit.bme.hu

Keywords: Future Internet, congestion control, complex networks, socio-economics

Future Internet research programs try to ignore and overcome the barriers of incremental development and encourage clean slate designs, propose new visions, architectures and paradigms for the coming 10-20 years. Recent results in congestion control research has shown that networks operating without explicit congestion control (like TCP) may survive without congestion collapse if appropriately designed in network resources and if end systems apply appropriate erasure coding schemes.
The exponential growth of the Internet makes it virtually impossible to manage the network with traditional centralized approaches (like the manager-agent one); hence research results of complex networks are expected to spread over the Internet with its autonomic behaviors. In this article we give overview and outlook of some interesting research areas connected to the future Internet research.

1. Introduction

The fast changes in the Internet usage and in its technology in the previous years have resulted in a growing expectation of possible paradigm changes in several mechanisms of the Internet. In this paper we address two challenging areas of these hot topics, namely, the solution for congestion problems and the management of future Internet.

The congestion is managed in the Internet by a congestion control mechanism called the Transmission Control Protocol (TCP), which is a complex transport protocol that has gone through several evolution steps since the beginning of the Internet. This evolution was driven by the ever-changing user and application requirements and also by the current technological limitations. As a result, a number of different TCP versions have been developed. TCP was originally designed as a reliable connection-oriented end-to-end transport protocol for the fixed wired Internet. However, the situation today is rather different concerning the pervasiveness of wireless technologies and also the increasing number of very high capacity links. It seems that the research community has arrived to a conclusion that it is very unlikely that an optimal TCP solution could ever be developed.

On the other hand, a new idea has arisen in the previous years, which basically says that we should design the solution for congestion problems from scratch. The idea challenges the researchers to think over the issue of congestion from a different point of view. What if we do not implement any congestion control in the network? How can the congestion be avoided in that case? Would it be possible to develop an Internet where the packet losses due to congestion events can efficiently be corrected by erasure coding? The TCP concept with its evolution and these exciting questions are discussed in the first part of the paper.

It is a well-accepted fact by current research concerning networks, that the future Internet will be characterized by the tons (in the order of trillions) of participating communicational entities and the complex and heterogeneous connections between them. Nevertheless the functional requirements of the next-generation networks will be highly diversified, and managing these networks with human interactions will be hardly solvable, thus the need for high-level automatic management is inevitable. Designing such complicated, large-scale systems is a complex task, and we still lack the adequate methodology for that. The means for describing large-scale networks developed a lot in the past years. We can see applications based on these findings, such as the search in complex networks, which will be discussed later in more detail.

Considering the future Internet, as in all complex systems, the analysis of the ongoing processes will require modern tools and methods. The conventional methodology, by which the global behavior of the system is treated as the collaborative functioning of different parts, lapses due to the complexity of the system. To handle these problems, we must come up with a self-organizing system model, where the centralized process control is replaced by distributed functioning and decentralized decision making. The monitoring of the network will take place on a macro level by analyzing the emergent features of the system, and not by independently observing the parts of it. In the second part of the paper, we discuss research topics in the area of large scale systems and consider self-organizing communication networks.

2. Congestion Control in Future Internet

Congestion control is a resource and traffic management mechanism to avoid and/or prevent excessive situations (buffer overflow, insufficient bandwidth) that can cause the network to collapse. It should not be confused with flow control, which prevents the sender from overwhelming the receiver. Congestion control has been accomplished by the *Transmission Control Protocol* (TCP) from the very beginning of computer networks and played an important role in the success of Internet.

The original protocol providing a reliable, connection-oriented service on top of IP networks dates back to 1981 (RFC 793). In the mid 1980s, serious incidents were experienced in the Internet when the network performance fell down by several orders of magnitudes. This phenomenon, called *congestion collapse,* raised the urgent need of some more sophisticated control mechanism in the transport layer. The original solution for the congestion collapse was provided in [1] by Van Jacobson. An essential part was added to TCP including the congestion control mechanisms. The congestion management of TCP is composed of two important algorithms. The *Slow-Start* and *Congestion Avoidance* algorithms allow the protocol to increase the sending data rate of sources without overwhelming the network and help to avoid congestion collapse. The protocol updates a variable called *congestion window* (`cwnd`, *w*) that directly affects the sending rate by means of limiting the number of unacknowledged packets in the network based on a sliding window mechanism which involves a *self-clocking* control. The congestion window variable is adjusted according to various algorithms in different phases of the connection. The basic mechanism was incrementally developed and tuned introducing new additional algorithms, e.g., RTO calculation and delayed ACK in 1989 (RFC 1122), SACK in 1996 (RFC 2018) and NewReno in 2004 (RFC 3782) just to mention a few. A standard TCP *(TCP Reno)* source starts sending according to the Slow-Start mechanism applying a multiplicative increase algorithm. More specifically, the congestion window is increased by a constant value for each acknowledgement received. This yields an exponential growth of the congestion window.

In Congestion Avoidance phase, the congestion window is adjusted by an AIMD (Additive Increase Multiplicative Decrease) mechanism which results in the classical “sawtooth” trajectory. When no packet loss is experienced, then the window is increased by $1/w$ per acknowledgements (AI) while it is halved as a response to packet loss (MD) as it is shown in the first row of *Table 2*. The main goal of the TCP’s congestion control mechanism is to provide good network utilization, to avoid congestion collapse and to share the resources (now the link capacities) among end-users in a fair way. This is achieved by a distributed, closed-loop feedback mechanism. The last requirement regarding the *fairness* properties of the protocol is an important part of the next-generation transport protocol design.

TCP congestion control had managed successfully the stability of the Internet in the past decades but it has reached its limitations in “challenging” network environments. The new challenges of next-generation networks (e.g., high speed communication or the communication over different media) generated an urgent need to further develop the congestion control of the current Internet. In recent years, several new proposals and modifications of the standard congestion control mechanism have been developed by different research groups all over the world. These new mechanisms and TCP versions address different aspects of future networks and applications and improve the performance of

Table 1. High speed transport protocols

Protocol	Type	Proposed by	Main properties
HighSpeed TCP	loss-based	S. Floyd, International Computer Science Institute (ICSI), Berkeley University of California, 2003.	AIMD
Scalable TCP	loss-based	T. Kelly, CERN & University of Cambridge, 2003.	MIMD
BIC TCP / CUBIC	loss-based	I. Rhee et al., Networking Research Lab, North Carolina State University, 2004/2005.	good utilization, stability, linear RTT-fairness
FAST TCP	delay-based	S. Low et al., Netlab, California Institute of Technology, 2004. (now: FastSoft Inc.)	promising fairness properties
TCP Westwood	measurement-based	M.Y. Sanadidi, M. Gerla et al., High Performance Internet Lab, Network Research Lab, University of California, Los Angeles (UCLA), 2001–2005	several versions, different estimation methods
Compound TCP	hybrid	K. Tan et al., Microsoft Research, 2005.	AIMD + delay-based component
XCP	explicit	D. Katabi et al., Massachusetts Institute of Technology (MIT), 2002.	modification of the routers is necessary

regular TCP. For example, standard TCP (Reno version) cannot provide acceptable performance in wireless or mobile environments where the propagation delay and the available bandwidth can suddenly change (e.g., during inter-system handover) which can result in multiple back-offs or in extreme cases in disconnection. In order to remedy this problem, new TCP versions have been dedicated to this environment. The drawbacks of standard TCP Reno can be experienced in high speed wide area networks, as well. These networks can be characterized by *high bandwidth-delay product* (BDP) and TCP cannot efficiently utilize them due to its conservative congestion control scheme. As a response to this problem, the research community has proposed several new transport protocols recently referred as *high speed TCPs* or *high speed transport protocols*.

The huge number of new ideas has resulted in different new TCP versions implemented in several environments. In order to select the "optimal" transport protocol, extensive performance analysis is necessary in a wide range of network environments and applications. In the recent years, many papers were published deepening our understanding of these new protocols regarding performance characteristics, co-existence issues, and other important properties affecting the possibilities of their deployment. In the rest of the paper, a brief overview is given on some promising TCP versions. The main properties of the most important variants are presented in *Table 1* while a more detailed overview can be found in [2].

In TCP Reno, the congestion event is indicated by packet losses and the sending rate is reduced when losses occur. Protocols considering this type of congestion measure are generally referred to as *loss-based* protocols. This one-bit congestion indication does not allow sophisticated congestion control mechanisms. In addition, the permanent oscillation which is an intrinsic property of this mechanism raises stability issues. Therefore, fundamentally different approaches have also been emerged. In the case of *delay-based* algorithms, the round-trip time (RTT) is regularly calculated during the connection and the sending rate is adjusted according to the current value of the average delay estimation. In other words, a "multi-bit" congestion measure (delay) is considered in the control decision. The most recent TCP versions combine both principles and apply *hybrid* or *combined delay/loss-based* mechanisms. Other solutions suggest explicit congestion feedback from the network routers. These mechanisms using *explicit congestion indication* require the modification of the routers, as well.

HighSpeed TCP (HSTCP) [3] is a modification to TCP's congestion control mechanism for use with TCP connections with large congestion windows. It changes the TCP response function to achieve better performance on high capacity links. HSTCP is based on an AIMD mechanism where the increase and decrease parameters ($a(w)$ and $b(w)$) are functions of the current value of the congestion window (see the corresponding row of Table 2) yielding an adaptive and more or less scalable algorithm. HSTCP introduces a new relation between the average congestion window and the steady-state packet drop (or marking) rate. It is designed to have the standard TCP response in environments with mild to heavy congestion (packet loss rates of at most 10^{-3}) and to have a different, more aggressive response in environments of very low congestion event rate.

Ideas to introduce MIMD mechanisms for TCP have also been considered. Scalable TCP (STCP) [4] is a good example which has been suggested as an efficient transport protocol for high speed networks. Here, the multiplicative increase and multiplicative decrease algorithm guarantees the scalability of the protocol. The congestion window is increased by a constant parameter (a) as a response to a received acknowledgement, while it is reduced in a multiplicative manner (by bw) in case of packet losses (see Table 2). A proposed setting for the constants are $a=0.01$ and $b=0.125$ [4].

Table 2. Details of some TCP versions

Protocol	Window adjustment	When	Reaction to loss
TCP Reno	$w \leftarrow w + \frac{1}{w}$	per-ACK	$w \leftarrow 0.5w$
HSTCP	$w \leftarrow w + \frac{a(w)}{w}$	per-ACK	$w \leftarrow w - b(w)w$
STCP	$w \leftarrow w + a$	per-ACK	$w \leftarrow w - bw$
BIC TCP	$w \leftarrow w + \frac{a}{w}, \quad a \in \left\{ S_{\min}, \frac{W_{\max} - w}{B}, \frac{w - W_{\max}}{B-1}, S_{\max} \right\}$	per-ACK	$w \leftarrow \beta w$
FAST TCP	$w \leftarrow \min\left\{ 2w, (1-\gamma)w + \gamma\left(\frac{baseRTT}{RTT} w + \alpha \right) \right\}$	periodically	$w \leftarrow 0.5w$

In order to solve the TCP's severe RTT (round-trip time) unfairness problems, BIC TCP has been developed [5]. BIC TCP combines two schemes called additive increase and binary search. When the BIC TCP source gets a packet loss event, the congestion window is reduced by a multiplicative factor (β); and the maximum window parameter (W_{max}) is set to the value of the congestion window just before the reduction while the minimum window parameter (W_{min}) is set to the current value. Then the protocol performs a binary search between these parameters by jumping to the "midpoint" between the bounds. (More exactly, this jump is based on the B parameter of the protocol.) If packet loss does not occur at the updated window size, that window size becomes the new minimum; if packet loss occurs, that window size becomes the new maximum.

An important restriction is also introduced, the growth cannot be more aggressive than a linear one with a constant parameter (S_{max}). This process continues until the window increment is less than a small constant (S_{min}), when the window is settles down around W_{max} (increasing slowly on a "plateau"). This mechanism yields an "AIMD-like" behavior where the growing function is most likely composed of a linear phase (additive increase) and a logarithmic one (binary search). When the updated window size exceeds the current maximum, then a new equilibrium state has to be found and BIC TCP enters into the max probing state. During this phase, the growing function is the inverse of the previous ones, more exactly, the window is increased exponentially first (which is very slow at the beginning) and then linearly.

This complex mechanism is also summarized in the corresponding row of Table 2. The good performance of the protocol, including good utilization, linear RTT fairness (RTT unfairness is proportional to the RTT ratio as in AIMD), good scalability, and TCP-friendliness, comes from the slow increase around W_{max} and the aggressive linear increase of additive increase and max probing phases. Further research with BIC TCP has been resulted in CUBIC. CUBIC [6] is an enhanced version of BIC TCP. It simplifies the BIC window control and improves its TCP-friendliness and RTT-fairness. It is worth noting that the default TCP protocol of Linux kernels from version 2.6.8 was BIC TCP. From kernel version 2.6.19, the default protocol is CUBIC.

The research on the delay-based ideas has resulted in FAST TCP [7]. FAST TCP has the same equilibrium properties as TCP Vegas but it can also achieve weighted proportional fairness. FAST TCP seeks to restrict the number of its packets queued through the network path between an upper (β) and a lower (α) bound, however, the behavior is usually controlled by a single parameter (α) that can be considered as the targeted backlog (packets in the buffers) along the flow's path [7]. Under normal network conditions, FAST TCP periodically updates its congestion window based on the comparison between the measured average RTT and the estimated round-trip propagation delay (when there is no queueing). More exactly, the window is adjusted according to the formula presented in Table 2, where γ is the step size affecting the responsiveness of the protocol, and `baseRTT` is the minimum RTT observed so far which is an estimation of the round-trip propagation delay. The parameter α controls the equilibrium behaviour, therefore the appropriate setting of this parameter is crucial. FAST TCP also reacts to packet losses by halving its congestion window.

The delay-based control also appears in other proposals like TCP-Africa [8]. TCP-Africa is a hybrid protocol that uses a delay metric to determine whether the bottleneck link is congested or not. In the absence of congestion it uses an aggressive, scalable congestion avoidance rule but in the presence of congestion it switches to the more conservative Reno congestion avoidance rule. The combination of the delay-based and the loss-based approaches also appears in TCP-Illinois [9]. TCP-Illinois uses loss as a primary congestion signal and delay as a secondary one. The protocol uses an AIMD mechanism but adjusts the increase and decrease parameters based on experienced queueing delay.

Compound TCP [10] is another important example where a synergy of delay-based and loss-based approach has been implemented. It uses a scalable delay-based component in the standard TCP Reno congestion avoidance algorithm. Compound TCP has been developed in the Microsoft Research and it is the default TCP protocol of Windows Vista and Windows Server 2008. Moreover, it can be installed for other Windows versions by downloadable hotfixes and in addition, the Linux implementation is also available.

The idea of incorporating accurate bandwidth estimations into the TCP congestion control has also opened a new path in TCP research. TCP-Westwood [11] is a prominent example where eligible rate estimation methods to intelligently set the congestion window and slow-start threshold have been introduced.

Another important group of congestion control protocols is based on explicit congestion notification instead of the implicit congestion signals such as packet loss or delay. These congestion control schemes require the assistance of network routers by this means the modification of the routers is also necessary. This is a serious disadvantage from the aspect of deployment feasibility. One of the main representatives of this group is the eXplicit Control Protocol (XCP) [12] which generalizes the Explicit Congestion Notification (ECN) proposal. Instead of the one bit congestion indication used by ECN, XCP capable routers inform the senders about the degree of the congestion at the bottleneck. In addition, XCP decouples the utilization control from fairness control.

The history of the research of congestion control protocols revealed that it is difficult to find an optimal protocol that meets all the challenges of the evolving Internet. It is very likely that the task to find a universal and optimal congestion control protocol is impossible. This view is supported by the fact that the developments of new applications show that they use their own congestion control mechanisms. These mechanisms in most of the cases are not *TCP friendly* so they cannot work together with TCP efficiently.

An interesting research is proposed in the framework of *GENI (Global Environment for Network Innovation)* advocating a *future internet without congestion control*. The basic idea is that flows do not attempt to relieve the network of congestion but rather send as fast as they can whenever they have data to send. Of course, if all flows are sending at maximal rates, then the packet loss rate within the network is probably high. To overcome this problem, flows can use efficient erasure coding.

This solution has several advantages but also raises some unsolved problems too. One of the biggest advantage is that we can achieve maximum resource utilization. It is because end hosts send packets as fast as possible and all available network resources between source and destination are utilized as much as possible. Links are constantly overdriven so any additional capacity is immediately consumed. This solution is *the most efficient regarding network resource utilization*. Another advantage is that this proposal can use *simple router architectures*. Routers no longer need to buffer packets to avoid packet loss so no need for expensive and power-hungry line-card memory. This can also result in significant decrease of the end-to-end packet delays for *supporting delay sensitive applications*. Moreover, this solution perfectly fits to a *network with all-optical cross-connects*.

With all these advantages we can also face a number of problems to solve. The most crucial question is that what *performance* can be achieved by implementing erasure coding techniques. The promising fact is that there is a number of new coding techniques proposed in the last decade with robust characteristics and high performance like fountain codes [13]. Another problem to solve is how to provide *fairness*. A mechanism is needed in the switches to perform selective packet dropping. As an example the *Approximate Fair Dropping* (AFD) [14] is a promising candidate to do this task.

Research is in the early phase but researchers can report some surprising results from their study. It seems that the congestion collapse is not as usual in networks without congestion control as it was believed [15]. Early results show that efficiency remains higher than 90% for most network topologies as long as maximum source rates are less than the link capacity by one or two orders of magnitude. It is also possible that a simple fair drop policy enforcing fair sharing at flow level is sufficient to guarantee 100% efficiency in all cases. Of course, there are several questions unanswered and new challenges to meet because present studies use some assumptions which are not fulfilled in practice.

An intensive research is needed to answer the exciting question whether we can build the future internet without congestion control and forget about TCP and its all problems.

3. Paradigm shift in managing networks

3.1 Large scale networks

One of the most important processes regarding the Internet today is the migration from the 20-year-old IPv4 protocol to the IPv6. The most demanding issue behind the process is that we are running out of the available IPv4 network addresses. The 32 byte address space of the IPv4 (~4.3 billion possible addresses) was created, when the computer (mainframe) to people ratio was 1:200. Nowadays this ratio is reaching 1:1 with 1.2 million users, which number can easily double in the near future due to the new users of the developing countries (according to estimates the number of users grow by 150 million every year). If we look at the spread of mobile devices, we can easily calculate that in the near future one person might even possess 200 network devices. According to assumptions, by 2010 the number of mobile devices will reach the number of PCs connected to the Internet. The typical tendency is that the PDA devices turn into a communicator device, but the real breakthrough will be the introduction of communications implants. In the world of sensors/actuators and intelligent materials we will see micro devices integrated into the human body, which are capable of wireless connectivity and will be used as life supporting or human-computer interaction devices. The RFID (Radio Frequency IDentification) is a really simple sensor network, where the active or passive devices can identify themselves, and it is already widely used.

The complexity of the Internet is growing not only by the addition of new network devices, but the rise of new available online services demands logically connected networks apart from the underlying physical one. This tendency is also reflected by the virtual ISPs (Internet Service Providers), mobile service providers and the development of virtual private networks. We can even mention the widespread use of peer-to-peer communicational methods, which all are built on the logically connected overlay networks. This significantly complicates the effective manageability of these networks.

For the effective analysis of large-scale complex networks first we need to develop such theory, which sets aside from the individual characteristics of the nodes, and concentrates on the structure of the connections and the character of the network. Moreover, the findings of this research field can be useful not only in the information technology. Many real world networks can be described with complex network models, for instance an organization, which is a network of people connected to each other. Also such networks are food webs, the global economical system or the connections between words in a language. We can also mention the diseases, which spread on the human social network (i.e. STDs). In general the research on complex networks concentrates on the various characteristics and the dynamic behavior of the networks.

Since the '50s the complex networks were described by the Erdős-Rényi [16] model, which was the only reasonable and adequately precise approach at that time. Still researchers presumed that real world networks are neither completely regular, nor completely random. The widespread use of computers and the Internet generated large databases, and these databases are easily accessible.

By analyzing these topological data, researchers made two important discoveries in the last two decades, one of them is the Watt and Strogatz "small world" effect; the other is the Barabási-Albert scale-free network model. The small world effect describes the same phenomena, which was presented in Milgram's famous experiment in the '60s [17]. He found that even if there are many billions of people in the world, the shortest route consists of only several hops between two randomly chosen individuals in the social network. The scale-free network model highlights another interesting feature of complex networks, namely the complex networks have scale-free degree distributions, and not Poisson distribution, which is the characteristic of random networks. The scale-free distribution means that it is highly probable for high degree nodes (hubs) to evolve in large-scale network *(Fig. 1)*.

There are three fundamental characteristic features of the complex network models, which are worth to emphasize: *average path length, clustering coefficient and degree distribution* [18]. Average path length is the average of the shortest distances of any two random nodes in the network. This distance represents the effective size of the network. It was an interesting discovery that most of the real world complex networks have relatively short average path length, which feature led to the name "small world". The examination of the basic parameters of complex networks was an important step in this scientific field. Based on these parameters, we can intuitively build up different mathematical models, which result in networks with similar statistical characteristics.

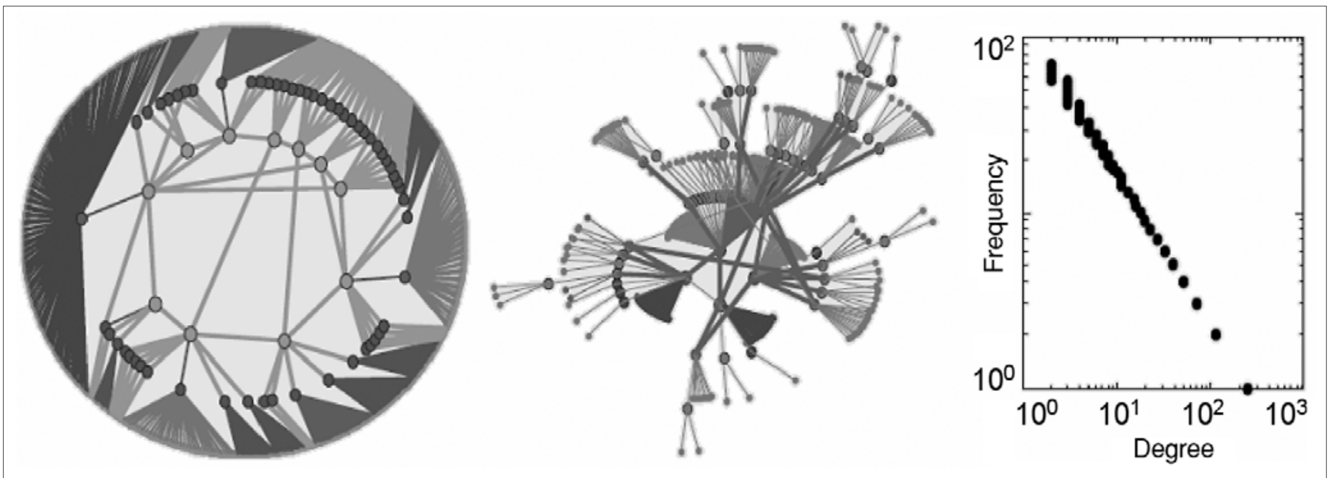
Another interesting topic in the field of complex networks is the problem raised by dynamic systems. Interesting observations were made about the synchronization of routing messages going over the Internet. Although the topology of the network was not specifically built for this purpose, the routers still synchronize their message exchanges easily, and when we break the synchronicity in one part of the network by introducing some randomness (altering a deterministic protocol), another part of the network will get in sync [19]. We can find more detailed treatment about these kinds of phenomena in the field of self-organizing systems.

3.2 Evolution and self-organization in communicational networks

Today telecommunication systems apply the widely used global network management paradigm. In this approach an external regulating unit controls the system. This unit constantly monitors the state of the system and the environment. When a problem occurs in the system, or the environment changes, the regulator calculates the appropriate response, and drives the system into the respective state *(Fig. 2)*. This solution is only feasible as long as we can find the appropriate response much faster, than the system changes its states. This criterion obviously delimits the permissible complexity and dynamism of the system, because the controller needs to be much more complicated than the system itself. For instance we can mention the link-state protocols, which cannot be used on large scale and complex topology.

A natural way to deal with complexity is self-organization, by utilizing the complexity of the system in the management plane. In a self-organizing system a large number of intricately connected devices achieve a global function by following simple local rules. It is shown in Fig. 2 that this way the control loop is integrated inside the system itself, thus the system can organize itself. Such a system evolves on its own obeying its given limitations. However, we don't know certainly what happens at a given point of the system, we can observe a precisely definable global behavior on the system level. Self-organization is not a feature of the system, but a paradigm, which can be helpful in understanding

Figure 1. Router level models of the Internet: engineering model (left), scale-free model (middle) and the degree distribution (right)



and designing certain real world systems (i.e.: complex telecommunication networks).

Algorithms showing classical signs of self-organization have played an important role from the beginning in the evolution and success of the Internet. We can mention the TCP protocol, which was previously discussed in the article. This protocol also uses a self-organizing technique while it deals with network congestion. It can control traffic flow parameters of a link in a decentralized manner, achieving an emergent result, such as fair resource distribution or high link efficiency. During the process every node makes strictly local decisions based on local information to reach a global goal.

Another example is the CSMA/CD algorithm, which is used in the Ethernet protocol. The rules of CSMA/CD guarantee that the communicating parties are able to detect the simultaneous transmissions and the consequent collisions on the common channel, and also prevent these collisions without a centralized control mechanism. If more nodes try to send packets on the same CSMA/CD channel at the same time, then the parties independently stop transmitting for a random time interval, hoping that next time they try to send packets, their packets won't collide with each other. If there is still collision on the channel, they increase the wait time interval, thus decreasing the probability of another collision. It is easily deductible that following these simple local rules the system realizes the fair and efficient resource distribution of the common transmission channel between the communicating parties.

The self-organizing systems belong to a highly active interdisciplinary research field, and they play an important role in many disciplines (biology, physics, social sciences). The systems utilizing these principles have many advantages, however we can only find few applications in the engineering fields. This can be explained by our lack of complete understanding concerning the mechanisms of self-organization. Designing such systems require an essentially new approach and design methods compared to our traditional ones.

3.3 Search in large-scale networks

Many large-scale networks, which are present in nature (human social networks, protein networks, neural networks, etc.), have good searchability as their important feature. Milgram's experiment [20] showed in 1961 that in human social networks not only short routes exist, but people are able to find them very efficiently depending on only the knowledge of a small local part of the whole network.

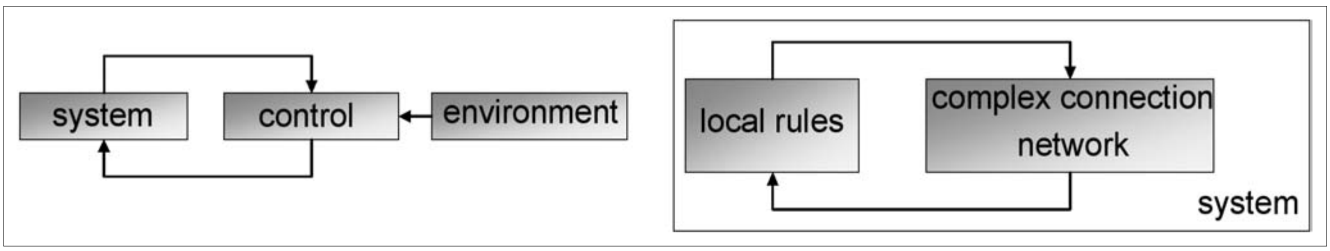
To achieve good searchability, which evolves in a self-organizing manner in nature, is a difficult task in artificial networks. An important example is any large-scale telecommunication network.

In the early development of the original concept of the Internet one of the most important elements was the design of a good routing protocol. The base of the framework that is still used consists of designated devices called routers, and the network emerging from the connections between these. This structure has the characteristic that some designated devices must have partial or even global knowledge of the whole topology to route messages. Since to gather the adequate information about the network topology takes time, the system must be quasi-static, and changes in the topology may happen with a limited speed. Considering that the next generation Internet will contain nodes by two or three times bigger order of magnitude than the present network, and these nodes will dynamically change the topology of the network, the original concept needs to be significantly adjusted.

The realization of efficient searchability is considered to be an important problem also in peer-to-peer networks. These networks utilize the Internet as a base structure, and they create an overlay network, which is responsible for the special P2P addressing and content searching functions. To cope with the strong network dynamism, and the high user activity as they connect and disconnect to the network, the system either uses a deterministic, but less scalable "network-flooding" technique, or apply a non-deterministic, more scalable method. These applications still highly depend on the underlying Internet infrastructure. Our research initiatives aim at developing an overlay network infrastructure (and the corresponding network protocols), which is able to handle numerous active (connecting, disconnecting or failing) users without a notable performance decline. The novel idea behind the research is the use of complex network structures instead of the ring structures present in P2P overlay networks.

The ongoing research about future networks more and more deals with "clean slate" designs. The recommended technologies show many similarities with the infrastructure-free ad-hoc networks. One of these concepts is the geographic position based addressing or geometry addressing and the search algorithms working with them. The communicating terminals can be mark-

Figure 2. The global management (left) and the structure of a self-organizing system (right)



ed by their geographical coordinates, and the routing is based on a simple greedy algorithm: if *X* is looking for *Y, X* first looks for its neighbor *Z*, which is closest to *Y*. In order for this algorithm to work, the network topology must meet some given conditions, for instance the Unit Disk Graph (UDG) is an applicable structure. In such cases, when the required connectivity conditions are not met, and for example it is true that *X* is not connected to *Y* and every neighbor of *X* is farther to *Y* thus there is no next step, some adjusted procedures need to be used in the algorithm.

The use of virtual coordinates can be the solution in this case. The task is to assign these virtual coordinates to the corresponding terminals, so that the greedy condition is realized, and the routing is always guaranteed [21]. The condition can be formalized in the following:

For every pair of nodes X and Y ($X \neq Y$) there exists a node Z that $d(Z,Y) < d(X,Y)$,
where $d(A,B)$ is the distance between A and B.

The virtual coordinates can be abstracted from the coordinates of the Euclidean plane or space, and they can be chosen from other abstract sets, after we defined a corresponding distance measure. The greedy routing capable virtual coordinate addressing assignment is called greedy embedding. If the greedy embedding is not possible due to the connection graph's special characteristics or other circumstances (i.e. dynamic topology change because of moving or failing nodes or links), we need to introduce complementary search procedures, for example face routing [22]. Our ongoing research deals with analyzing complex network search algorithms and topology management algorithms, which are efficient at given topology constraints (i.e. max node degree).

The routing techniques presented above have a common feature: they are easily scalable for large-scale networks, because each node needs to have information only about its neighbors. There is no need for large routing address tables, and the decision-making mechanism is really simple. This type of techniques is often called router-free routing. The procedures can be really effective, if they are supplemented with addressing algorithms, which can provide network addresses in a self-organizing manner based on local rules. Our research aims at realizing applications in accordance with our theoretical findings concerning these issues.

The management of large-scale, dynamic and structure-free networks is an active research field, and although we already have many basic results at will, the great challenges of technological applications are still ahead of us.

4. Conclusion

In this paper an overview was given about two research fields of the future Internet research where paradigm changes are expected. In the first part the issue of Internet congestion control was discussed. It was presented that Transmission Control Protocol (TCP) has always been serving as a solution for handling and avoiding congestion problems in the Internet. The mechanism of TCP was discussed and a short insight was given into the TCP versions that were developed during the history of the Internet.

A new idea is also discussed for solving the congestion problems in the future Internet. It is not based on control but rather on erasure coding. This solution makes the tempting promise that a future Internet could be developed without any congestion control. The possibility of this solution is a topic of current research. In the second part of the paper the manageability issues of large scale complex networks have been discussed. It is shown that in the future Internet network management methodologies featuring self-organization must play an important role. As a widely researched area, the special topic of searching in large networks has been presented in more detail.

Authors

MÁRTON CSERNAI is an undergratuate student on the Budapest University of Technology. He is about to finish his MSc degree in next-generation communication networks. He is participating in the ongoing research on future internet technologies at the Department of Telecommunications and Media Informatics. His main research fields are large scale complex networks and next-generation communication networks.

ANDRÁS GULYÁS received M.Sc. and Ph.D. degree in Informatics at Budapest University of Technology and Economics, Budapest, Hungary in 2002 and 2008 respectively. Currently he is a research fellow at the Department of Telecommunications and Media Informatics. His research interests are complex and self-organizing networks, network calculus and traffic management.

ZALÁN HESZBERGER recieved his M.Sc. and Ph.D. degree in electrical engineering at the Budapest University of Technology and Economics (BME), Budapest, Hungary in 1997 and 2007, respectively. Currently he is an assistant professor at the Dept. of Telecommunications and Media Informatics at BME. His main research interests are future internet technologies and complex networking.

SÁNDOR MOLNÁR received his M.Sc. and Ph.D. in electrical engineering from the Budapest University of Technology and Economics (BME), Budapest, Hungary, in 1991 and 1996, respectively. In 1995 he joined the Department of Telecommunications and Media Informatics, BME. He is now an Associate Professor and the principal investigator of the teletraffic research program of the High Speed Networks Laboratory. Dr. Molnár has been participating in several European COST and CELTIC research projects. He served on numerous technical program committees of IEEE, ITC and IFIP conferences. He is active as a guest editor of several international journals and is serving in the Editorial Board of the Springer Telecommunication Systems journal. Dr. Molnár has more than 130 publications in international journals and conferences. His main interests include teletraffic analysis and performance evaluation of modern communication networks.

BALÁZS SONKOLY received his MSc degree in software engineering from the Budapest University of Technology and Economics in 2002. Now he is a research engineer at the Department of Telecommunications and Media Informatics. His interests are in the field of traffic modeling and high speed transport protocols. He has authored international journal and conference papers in the are of high speed networks and the evaluation of TCP protocols.

References

[1] V. Jacobson, Congestion avoidance and control, In Proceedings of ACM SIGCOMM '88, pp.314–329., Stanford, CA, USA, 16-18 August 1988.

[2] S. Molnár, B. Sonkoly, T.A. Trinh, A Comprehensive TCP Fairness Analysis in High Speed Networks, Computer Communications, Elsevier, Vol. 32, Issues 13-14, pp.1460–1484., August 2009.

[3] S. Floyd, Highspeed TCP for large congestion window, IETF RFC 3649, December 2003.

[4] T. Kelly, Scalable TCP: Improving performance in high speed wide area networks, ACM SIGCOMM Computer Communication Review, 33(2):83–91, April 2003.

[5] L. Xu, K. Harfoush, I. Rhee, Binary increase congestion control (BIC) for fast long-distance networks, In Proceedings of IEEE Infocom '04, Vol. 4, pp.2514–2524., Hong Kong, China, 7-11 March 2004.

[6] I. Rhee, L. Xu, CUBIC: a new TCP-friendly high-speed TCP variant, In Proceedings of Third International Workshop on Protocols for Fast Long-Distance Networks (PFLDnet 2005), Lyon, France, 3-4 Februar 2005.

[7] D.X. Wei, C. Jin, S.H. Low, S. Hegde, FAST TCP: motivation, architecture, algorithms, performance, IEEE/ACM Transactions on Networking (ToN), 14(6):1246–1259, 2006.

[8] R. King, R. Riedi, R. Baraniuk, Evaluating and improving TCP-Africa: an adaptive and fair rapid increase rule for scalable TCP, In Proceedings of Third International Workshop on Protocols for Fast Long-Distance Networks (PFLDnet 2005), Lyon, France, 3-4 Februar 2005.

[9] S. Liu, T. Basar, R. Srikant, TCP-Illinois: A loss and delay-based congestion control algorithm for high-speed networks, In Proceedings of First International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS), Pisa, Italy, 11-13 October 2006.

[10] K. Tan, J. Song, Q. Zhang, M. Sridharan, A compound TCP approach for high-speed and long distance networks, In Proceedings of IEEE Infocom '06, Barcelona, Spain, 23-29 April 2006.

[11] R. Wang, K. Yamada, M.Y. Sanadidi, M. Gerla, TCP with sender-side intelligence to handle dynamic, large, leaky pipes, IEEE Journal on Selected Areas in Communications 23(2):235–248, 2005.

[12] D. Katabi, M. Handley, C. Rohrs, Congestion control for high bandwidth-delay product networks, In Proceedings of ACM SIGCOMM '02, Pittsburgh, PA, USA, 19-23 August 2002.

[13] M. Luby, LT-codes, The 43rd Annual IEEE Symposium on the Foundations of Computer Science, pp.271–280., 2002.

[14] R. Pan, L. Breslau, B. Prabhakar, S. Shenker, Approximate fairness through differential dropping, ACM SIGCOMM Computer Communication Review, Vol. 33, Issue 2, April 2003.

[15] T. Bonald, M. Feuillet, A. Proutière, Is the "Law of the Jungle" sustainable for the Internet?, IEEE INFOCOM '09, Rio de Janeiro, Brazil, 19-25 April 2009.

[16] P. Erdős, A. Rényi, "On the evolution of random graphs", Publ.: Math. Inst. Hung. Acad. Sci., Vol. 5, pp.17–60., 1959.

[17] S. Milgram, "The small-world problem", Psychology Today, Vol. 2, pp.60–67., 1967.

[18] Xiao Fan Wang, Guanrong Chen, Circuits and Systems Magazine, IEEE, Vol. 3, Issue 1, pp.6–20., 2003.

[19] S. Floyd, V. Jacobson, "The synchronization of periodic routing messages," IEEE/ACM Trans. Networking, Vol. 2, No. 2, pp.122–136., April 1994.

[20] Milgram, Stanley, "Behavioral Study of Obedience", Journal of Abnormal and Social Psychology, 67(1963):371–378.

[21] Cedric Westphal, Guanhong Pei, Scalable Routing Via Greedy Embedding, In Proceedings of IEEE INFOCOM'09 Mini-Conference, Rio de Janeiro, Brazil, April 2009.

[22] J. Li, L. Gewali, H. Selvaraj, V. Muthukumar, "Hybrid Greedy/Face Routing for Ad-Hoc Sensor Network," Euromicro Symposium on Digital System Design (DSD'04), pp.574–578., 2004.