Network Operator Independent Resilient Overlay for Mission Critical Applications (ROMCA)

Xian Zhang Chris Phillips
Networks Research Group, Department of Electronic Engineering and Computer Science
Queen Mary, University of London
London, United Kingdom
xian.zhang@elec.qmul.ac.uk, chris.phillips@elec.qmul.ac.uk

***Abstract*—This paper proposes a Resilient Overlay for Mission Critical Applications (ROMCA); a novel operator-independent overlay architecture providing a resilient and reliable service across wide-area networks. One feature of ROMCA is that its overlay topology can be altered according to the underlying network conditions. Moreover, resilience is achieved by combining centralized topology construction control and distributed dynamic mapping of paths onto the overlay topology according to network conditions. ROMCA can mitigate the shortcomings of the Internet network infrastructure and provide low recovery times in the event of network failure(s).**

Keywords-ROMCA; resilience; reliability; mission critical

I. INTRODUCTION

Mission critical applications, e.g. remote control messaging, require high reliability and are very sensitive to network failure(s) and performance degradation, so there is considerable interest in developing a cost-effective scheme to provide customers with a resilient delivery service. Currently the Internet only provides “best effort” packet transport. Furthermore, Border Gateway Protocol (BGP), used for routing across Autonomous Systems (AS) is characterized by re-convergence times of several minutes or longer [1]. These factors limit the ability to support applications with higher reliability requirements; thus necessitating the exploitation of new strategies.

Inspired by the research demonstrating that overlay networks can be an effective means of circumventing the shortcomings of the Internet by supporting new applications without introducing changes to the existing network infrastructure [2, 3, 4, 5, 6, 7, 8 and 9], a novel architecture called Resilient Overlay for Mission Critical Applications (ROMCA) is proposed. This permits mission critical services to be set up and maintained despite uncertainty in underlying wide-area networks.

The rest of the paper is structured as follows. Firstly, in Section II, we present a state-of-the-art review. Then, in Section III and IV, our overlay architecture is illustrated and explained in detail together with operational examples. Moreover, Section V discusses the ROMCA topology construction process based on simulation and draws conclusions that can be used as the guideline for ROMCA topology construction. Finally, conclusions and ongoing work are briefly given in Section VI.

II. BACKGROUND

An Overlay Network (ON) is generally composed of selected nodes from the underlying network. By monitoring the performance of virtual links between these overlay nodes and taking advantage of the Internet redundancy in case of failure(s), the ON can support new applications with more stringent QoS (Quality of Service) requirements. For instance, Tapestry [3] and CAM-Chord [4] deploy an overlay layer on top of the Internet to support peer-to-peer (P2P) and multicast applications, respectively.

In this paper, our targeted service is to support mission critical applications, which usually require high network availability, i.e. to make network failures(s) or performance degradation not easily perceived from the users’ perspective with the help of auxiliary strategies. More specifically, additional mechanisms are deployed based on the Internet to meet the QoS and network resilience requirements of this kind of service. Till now, several principle architectures have been proposed aimed at improving Internet service performance in terms of resilience, QoS such as latency, packet loss ratio and so forth.

RON [2] is the first proposed overlay network aimed at improving the resilience of the Internet. The nodes in RON form a full mesh topology and use both active probing and passive measurements to monitor the Internet performance. If a working path undergoes failure(s) or it can no longer satisfy the prescribed performance requirements, the traffic is diverted to an intermediate RON node bypassing the unsatisfactory path segment. In contrast to the high convergence time of BGP, RON can achieve recovery times in the order of tens of seconds based on test-bed experiments. However, RON is application-specific and the number of overlay nodes is assumed to be no more than 50. It could incur scalability problems when widely deployed as RON utilizes a fully meshed topology and active probing and monitoring. The scalability issues of RON have been considered in [8], where hierarchically organized link state routing is employed.

The second type of overlay architecture is one that is provider-dependent, whilst attempting to fulfil the QoS requirements of its customers. For example, the objective of SON [5] is to find overlay paths under certain bandwidth constraints. Another example is QRON [6]. It utilizes a kind of overlay node called Overlay Broker to construct a hierarchical topology, aimed at finding QoS-satisfied paths for end users.

The nodes in QRON subscribe to an Internet Service Provider (ISP) for high bandwidth connections, thus the primary concern is how to provide the service in a cost-effective way.

A third type of architecture is exemplified by OHSR [12]. This is a simple, scalable method for improving resilience, proposed after extensive measurement research on the characteristics of the Internet path failure(s). Nodes in OHSR will not form an overlay but try to recover from path failures by routing indirectly using randomly chosen intermediaries without path monitoring. Exploiting the idea of OHSR, HORNS [13] proposes a heuristic node selection algorithm to support interactive real-time applications. The difference between the two is that HORNS improves OHSR performance by keeping the candidate node pool small and using end-to-end (ETE) delay as the selection criterion for the intermediary candidate.

ROMCA shares the same objectives with these architectures, namely enhancing the performance of the Internet, especially in supporting of applications with stringent requirements. However, its characteristics differ in regard to the following features:

- **Network provider independence:** Based on the multi-domain Internet, ROMCA is designed to be a network-provider-independent overlay architecture, which can ensure the flexibility of its deployment across the wide area networks and avoid issues of inter-provider trust.
- **Scalable overlay topology:** In contrast to RON, the overlay nodes in ROMCA form into a partially meshed topology with layer-3 diversity. Moreover, it can be altered dynamically according to the results collected by exploiting ICMP-based traceroute methods. Thus, it can provide effective virtual linking between two overlay nodes that may be situated in different domains and perform path selection based upon customer service requests.
- **High Resilience:** Although, ROMCA shares the same objective of promoting the resilience of the Internet using alternative node(s) as RON and OHSR do, it employs a scalable topology that can guarantee the resilience of working and backup paths to some extent. Moreover, ROMCA employs well-researched protection/restoration methods used in connection-oriented networks, which can provide the customer with highly resilient paths based on the dynamic overlay topology.

III. ROMCA ARCHITECTURE

As shown in Figure 1, ROMCA consists of an Overlay Directory Service (ODS) and Overlay Gateways (OG) chosen from different ASes.

The ODS is a centralized component responsible for service access and managing the overlay topology, including not only the acceptance and removal of overlay nodes, but also selection of the OG adjacencies. The ODS thus determines which virtual links will be set up between the OGs. However, it plays no part in the actual forwarding of traffic across the overlay. OG nodes are responsible for service data forwarding (see [14] for further functional details).

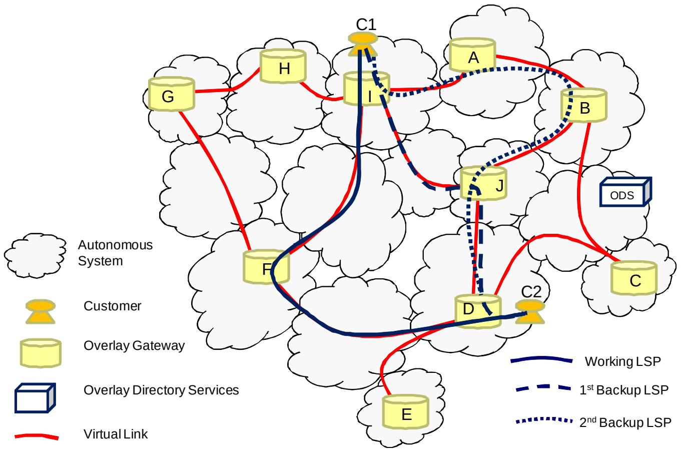

Figure 1 ROMCA Architecture

As [9] concludes, the overlay topology has an impact on overlay performance and physical network information is helpful in constructing the overlay topology. A moderate amount of work has been dedicated to discussing the overlay topology building process. One example is SIMON [7], which employs a hierarchical distributed server mechanism to organize the intra-domain and inter-domain overlay nodes. We use a different method in ROMCA, namely employing the ODS for organizing the overlay topology taking into account the position of potential overlay nodes. Simulation analysis described in Section IV provides some basic principles for forming the overlay topology. In terms of the IP path and topology discovery mechanism, [10] compares different traceroute methods and examines their performance. For example, ICMP-based traceroute tends to reach more destinations and collect “presence” information of a greater number of AS links as compared with other methods. Indeed, [11] confirms the superiority of traceroute method for discovering the Internet topology over alternative discovery methods. Thus, ICMP-based traceroute methods are used in supporting the construction of a layer-3 diversified overlay topology and are also used for performance monitoring of the virtual links between the OGs.

Together, the single ODS and multiple OG entities form the ROMCA architecture and are the means by which ROMCA provides resilience service to end-users. The architecture itself is effectively hidden from the end-users, which simply know the public address of the ODS from which the appropriate OG points-of-presence can be ascertained.

IV. OPERATIONAL EXAMPLES

This section first provides operational examples together with performance information. Furthermore, conclusions drawn from simulations are made in relation to the overlay topology construction process.

A. Topology Construction

It can be seen from the Figure 1 that the overlay topology is partially meshed and generally organized into inter-connected cycles, though stub connections are permitted. The virtual links between adjacent OGs nodes are chosen according to probing results and network performance measures. For instance, assume node G applies to the ODS to join the overlay. After retrieving the potential neighbouring information from the ODS and trace-routing to these corresponding nodes, G reports its findings to the ODS and is accepted into the ROMCA topology as it is a "valuable" transit node having Layer-3 diversified paths to H and F. There are also situations where stub nodes may get accepted according to their resilience and service access utility. For example, node C is viewed as a potential alternative path for B and D, so there are two virtual links connected from C. Whereas, the stub node E is only accepted into the overlay by establishing one virtual link connected to D. Node E is used as a service access point for the customer in its AS.

An example of the basic joining process for new nodes is depicted in Figure 2 (i.e. for node E) using a UML sequence diagram. As explained previously, the overlay topology is strategically constructed and maintained by the ODS, but a distributed mechanism for topology updating and network performance information flooding is needed so that OGs can maintain up-to-date performance information to enable them to efficiently establish working and backup paths for customer traffic. In our architecture, a flooding mechanism similar to that used in Open Shortest Path First (OSPF) and constraint-based routing protocols are deployed among the OGs. So, when the performance of a virtual link changes across a threshold, update packets will be flooded to all OGs so that they can store the updated information in their Link-State Database (LSD). This in turn will influence the routes chosen by the ingress OG for subsequent working and backup paths.

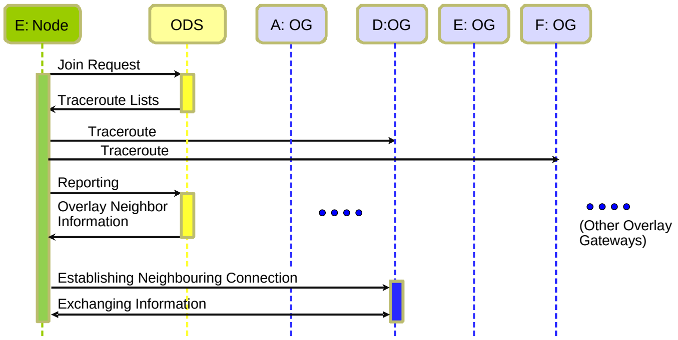

Figure 2 OG Node Joining Process

B. *Service Provisioning Example*

To explain the service provisioning operation, consider the topology depicted in Figure 1, where Multi-Protocol Label Switching (MPLS) is used to define the working and protecting Label Switched Paths (LSPs). Consider a ROMCA customer (i.e. customer 1) located in the same Domain as OG I. The customer approaches the ODS providing the IP address of itself and the IP address of the destination customer (i.e. customer 2) from which the ODS can infer their proximity to the various OG nodes that are operational. In this case, the ODS knows the customer has no desire to become an OG; it simply wishes to exploit the overlay mesh to provide the resilient pathways to a fellow customer in another AS. The ODS provides it with the nearest point-of-presence, i.e. the address of OG I, giving it a connection "ticket". The ODS also informs the local OG, i.e. I, that it can expect an approach from customer A and the customer's needs as well.

When customer 1 contacts OG I, I checks the ticket details. In this case, the customer wishes to establish disjoint resilient paths to fellow customer 2. Using its LSD, OG I sends RSVP-TE path messages to OG D using as diverse links and nodes where possible. For example the working path may be taken to be via OG I, F and D. The corresponding protection path could be: OG I, J and D. Once established, a FEC-to-Label binding entry is created at OG I, and customer 1 is informed that the service is ready. Traffic from customer 1 to 2 now uses IP to reach OG I (tunnelling IP in IP). There the packet will be encapsulated using the format depicted in Figure 3.

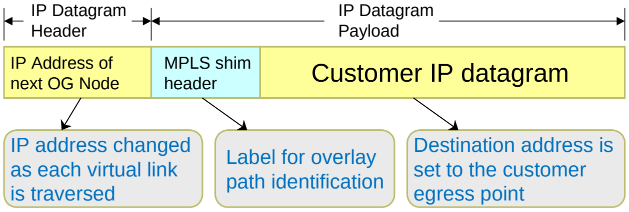

Figure 3 Packet Format used between OGs

Accordingly, the packet has an MPLS label pushed onto it and this is encapsulated in a datagram for OG F. At intermediate OGs, the MPLS shim layer is examined and the label is swapped and the traffic re-encapsulated and sent to the next-hop OG, and so on. At node D, the label is popped and the datagram delivered to customer 2 as per standard IP.

C. *Resilience Implementation Scenarios*

If a failure happens in a transit AS, it may take the ASes several minutes to re-converge and thereafter find the proper route to divert the traffic accordingly. But in ROMCA, as the neighbouring OGs exchange "hello" messages periodically (e.g. several seconds), the failure will result in loss of these heartbeat messages. Once the time threshold for neighbouring connectivity loss is reached, the OG(s) adjacent to the point(s) of failure will propagate the information over the virtual links to the ingress point(s), which can immediately update the FEC-to-Label binding so that the traffic is mapped onto the pre-configured diversified protection path(s). These paths avoid the failed AS and so can ensure that service delivery is quickly re-established.

As the service provisioning example shows, when ROMCA is used, traffic from customer 1 to 2 goes via the ingress OG I, and from there, follows a working path dictated by the LSP. Moreover, a protection path is also set up for resilience purposes. In the event of a failure in an AS lying between OG I and F, the ingress OG will switch the traffic to the protection LSP, i.e. to the path going from OG I, J to D, thus re-establish customer data packets transmission typically within seconds.

Another example is the dynamic mapping of LSPs according to the updated monitoring results. If the virtual link between I and J results in a longer delay than that of the path from I, via A, B to J, the backup LSP can be dynamically changed to the alternative backup LSP depicted in Figure 1, while the working LSP remains unchanged.

D. Performance Considerations

- Dynamic Overlay Topology

One feature of the ROMCA architecture is the adoption of a centralized means of topology construction, while providing services for customers in a distributed manner. The topology building strategy, utilizing the ICMP-based traceroute method among OGs, can ensure the creation of an efficient overlay topology and layer-3 diversity of virtual links to some extent, thus facilitating the resilience mechanisms deployed in the overlay. Moreover, as the performance and connectivity of the underlying paths change, the topology can be altered accordingly. It is also able to operate in a sparse deployment environment where multiple ASes may exist between adjacent Overlay Gateways, as is shown in Figure 1.

- High Resilience

Another feature of ROMCA is that both reactive and proactive methods are exploited to meet the high resilience demands of service customers. For instance, well-established protection mechanisms, such as 1:1, 1+1 and p-cycle, adopted in connection-oriented networks, can be employed. What's more, machine learning can be incorporated to take advantage of the historical network performance data and make changes of working and backup paths before failure(s) occur. In the short term, failure(s) can be detected using regular "hello" messages exchanged between adjacent OGs; in the long term, we expect actions will be taken using prediction before outages interfere with customer services, by altering the virtual link arrangement.

- Low Recovery Time

Given each OG exchanges regular neighbour heartbeat messages (i.e. hellos) and the protection/restoration method adopted in overlay, we believe ROMCA can mitigate the slow convergence characteristic of BGP and achieve lower recovery time in the order of tens of seconds or even seconds on average. The absence of hellos along a virtual link triggers the switchover to backup paths (typically from the ingress OG), irrespective of the information being disseminated between ASes by BGP.

V. TOPOLOGY CONSTRUCTION, SIMULATION AND ANALYSIS

A. Problem Formulation

The impact of overlay node degree and underlying network topology information availability on the ROMCA topology construction is investigated using simulations. The topology construction algorithms proposed in this paper employ the procedure flow depicted in Figure 2.

The ROMCA overlay topology construction process can be formulated as follows:

Input:

- Physical Network Gp(Vp, Ep);
- Overlay Node Number, No;
- Maximum node degree (ND) requirement based on the conclusion drawn on Part B of this section;

Output: Overlay Network Topology Go(Vo, Eo);

Objective: Maximize overlay network virtual link diversity so that it can maintain the same network performance with lower overhead;

B. Node Degree Impact

The authors in [15] have observed that overlay node degree above certain threshold will gain little in improving overlay failure recovery ratio whereas incur larger amount of routing and monitoring overhead in the overlay layer. So simulations are carried out to determine an appropriate overlay node degree used for ROMCA overlay topology construction. The node degree simulations are carried out using two underlying topologies: a random topology and a scale free topology generated by Pajek [16]. The former is composed of 50 nodes and 100 links, while the latter consists of 50 nodes and 130 links. The failure distribution is uniformly distributed among all physical links. Overlay nodes are fixed to 15 and Failure Recovery Ratio (FRR), which means the percentage of recovered overlay paths in all the affected overlay paths in the overlay defined in [15], is used as the performance criteria.

The two simulations based on different topologies show similar trend. Due to space limitations, only the one for scale free network topology is shown in Figure 4.

In the figure, FRR Differentiation Factor Δ(k), as shown in Equation 1, is used to represent the difference between FRR of full mesh scheme with that of other node degree schemes, where k stands for the node degree constraint used in topology construction process.

$$\Delta(k) = |\,FFR(FM) - FFR(k)| \qquad (1)$$

From the figure, similar conclusions can be drawn that for the node degree after a certain value, the network failure recovery ratio will have a diminishing improvement, but the overlay overhead goes up with the increasing node degree. Therefore, in subsection C, 6 is set as the maximum node degree for the simulations.

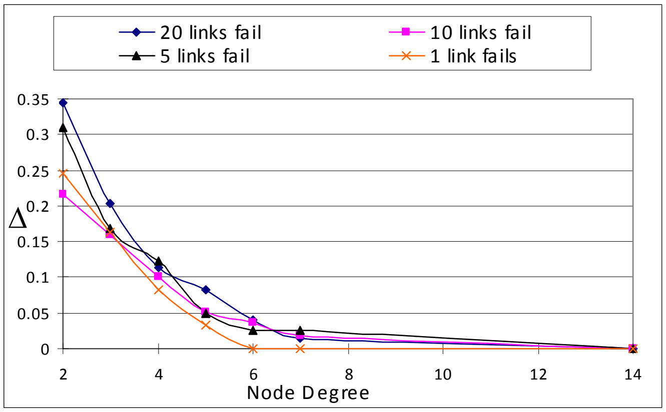

Figure 4 Node Degree Impact on Overlay Topology Construction

C. *Topology Construction Simulation*

It can be seen from Figure 2 that both the overlay nodes and the virtual links in ROMCA are not pre-determined. In this paper, the physical topology information is assumed to be available for the overlay topology construction process. Moreover, the overlay nodes are randomly selected with only maximum overlay node constraint and virtual links are set up using the following two algorithms:

- K-Most Diversified First (K-MDF): under the maximum node degree constraint, the new joining node will calculate the diversity degree of all potential neighbour OGs and choose the K number of OGs with the highest diversity degrees;
- K-Most Diversified First with Discarding (K-MDFD): this algorithm is the same as the first one with the exception that it will not choose an OG node with 0 diversity degree as its new neighbour. Thus it can further prune links that are physically co-located with the existing overlay links.

In the simulation, diversity degree is defined as the percentage of new physical links in the virtual overlay link that will be set up. Simulations are based on the scale free topology used in the previous subsection. For comparison purpose, Full Mesh (FM) and K Random Connection (KRC) are included.

The Failure Detection Ratio (FDR) and actual average ND obtained from each schemes under 10 IP link failures are compared and shown in Figure 5. For other link failures, the trend is similar. From the simulation results it can be inferred that the proposed algorithms can maintain the same performance while lowering the overlay overhead by reducing its node degree. The simulations prove that underlying physical path information is helpful in constructing an efficient ROMCA overlay topology. These observations are now being used as an input in our further work undertaking a more in-depth performance analysis of ROMCA.

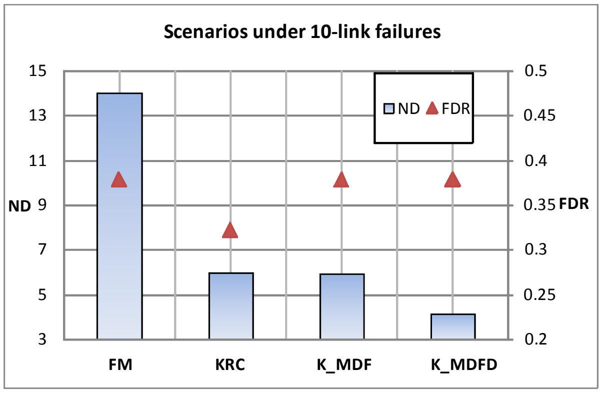

Figure 5 ROMCA Topology construction simulation results

VI. CONCLUSIONS AND ONGOING WORK

In this paper, a novel overlay architecture named ROMCA for supporting the mission critical applications is presented. Its mechanisms are explained, with which the architecture can provide resilience with low recovery times in response to network failure(s). In addition, ROMCA requires no specific support from network operators. This enables the overlay to be offered as a value-added enterprise service that can be deployed incrementally. It also capitalizes on the wealth of knowledge developed for resilience in existing circuit switched networks. Some conclusions are presented as the guideline for building an efficient overlay topology.

Simulations about the ROMCA topology construction are carried out and initial results given show (1) selecting an appropriate node degree is useful in balancing network failure recovery ratio and overlay overhead, and (2) that the underlying physical path information is instrumental in overlay topology construction. Ongoing research is now evaluating its performance in terms of failure recovery ratio, overlay routing overhead and prediction performance both using a simulation platform and in field experiments.

REFERENCES

[1] Craig Labovitz, et al. “Delayed Internet routing convergence”, *IEEE/ACM Trans. Networking*, Vol.9, No.3, pp. 293-306, June 2001.

[2] David G. Anderson, et al. “Resilient Overlay Networks”, *In Symposium on Operating Systems Principles*, pp. 131-145, 2001.

[3] Ben Y. Zhao, et al. “Tapestry: a resilient global-scale overlay for service deployment”, *IEEE Journal on Selected Areas in Communications*, Vol. 22, Issue 1, pp. 41 – 53, Jan. 2004.

[4] Zhan Zhang, et al. “Resilient Capacity-Aware Multicast Based on Overlay Networks”, *IEEE International Conference on Distributed Computing Systems (ICDCS)*, pp. 565 – 574, 2005.

[5] Zhenhai Duan, Zhi-Li Zhang, et al. “Service overlay networks: SLAs, QoS, and bandwidth provisioning”, *IEEE/ACM Transactions on Networking*, Volume 11, Issue 6, pp. 870 – 883, Dec. 2003;

[6] Zhi Li, Prasant Mohpatra. “QRON: QoS-aware routing in overlay networks”, *IEEE Journal on Selected Areas in Communications*, Vol. 22, Issue 1, pp. 29 – 40, Jan. 2004.

[7] Elaoud, M. McAuley, G. Kim, J. Chennikara-Varghese. “Self-initiated and self-maintained overlay networks (SIMONS) for enhancing military network capabilities”, *Military Communications Conference, 2005. MILCOM 2005. IEEE*, Vol. 2. pp. 1147- 1151, 17-20 Oct. 2005.

[8] Hasegawa, G., et al. “Emergency Communication Services Based on Overlay Networking Technologies”, *Fourth International Conference on Networking and Services, ICNS 2008*, 2008.

[9] Li Zhi, Prasant Mohapatra. “The impact of topology on overlay routing service” *Proceedings of IEEE INFOCOM 2004 - Conference on Computer Communications*, pp. 408-418, 2004.

[10] Matthew Luckie et al. “Traceroute Probe Method and Forward IP Path Inference”, *IMC’ 08*, pp. 311-323, October 20-22, 2008.

[11] Bin Yuan, Guoqiang Zhang, Yanjun Li, Guoqing Zhang, Zhongcheng Li. “Improving Chinese Internet’s Resilience through Degree Rank Based Overlay Relays Placement”, *ICC 2008*, pp. 5823-5827, 2008.

[12] Gummadi, K.P., et al. “Improving the reliability of internet paths with one-hop source routing”, *6th conference on Symposium on Operating Systems Design & Implementation*, Vol. 6. 2004.

[13] Yin, C. et al. “Heuristic Relay Node Selection Algorithm for One-Hop Overlay Routing”, *in Distributed Computing Systems Workshops (ICDCS)*, 2008.

[14] Xian Zhang, Chris Phillips, “Network Operator Independent Resilient Overlay for Mission Critical Applications”, *the annual PostGraduate Network Symposium (PGNet)*, accepted for publication, 2009.

[15] Zhi Li, Prasant Mohapatra, “On Investigating Overlay Service Topologies”, Computer Networks, pages 54-68, vol.51, 2007.

[16] http://vlado.fmf.uni-lj.si/pub/networks/pajek;