Research Statement

Michalis Faloutsos

My research focuses on computer security and networks and can be grouped into the following, partially-overlapping areas:

1. **Web security** focusing on the detection of malware on websites, and the identification of the mechanisms deployed by hackers and spammers.
2. **Network science, complex systems and social networks**, where the goal is to model the topology, understand the evolution, identify patterns, and detect anomalies, including physical network topologies, communication interaction, biological networks, and human-centric networks like eBay and youtube.
3. **Network security and measurements** focusing on classifying traffic and detecting malware and intrusions in enterprise and backbone networks.
4. **Network protocols** focusing on improving the performance of protocols in the areas of: Internet routing (BGP), robust and scalable ad hoc routing, contend distribution, privacy and anonymity, and group communications.

My philosophy. The driving forces for my work aspire to be: (a) practical impact, (b) contributions to fundamental science, and (c) enabling students to achieve their career goals. As we all know, reality is not as clean-cut as this. However, my publication and funding records provide indications of some successes towards these goals.

Regarding the third goal, I find that it is pretty much taken care of by fulfilling the first two goals: the challenge is more the human-resource management. The key is to unleash the true potential of each student by guiding their abilities to topics according to their preference, strength, and long term goals. Over the years, I have had the pleasure of working with excellent students, who arguably taught me more than I taught them, and got prestigious positions at academic institutions (e.g. J. Eriksson at U. of Illinois, Chicago, K. Pelechrinis[1] at U. Pittsburgh), research labs (e.g. T. Karagiannis at Microsoft Research, G. Siganos at Telefonica Research) and industry (e.g. Y. He at Yahoo, I. Broustis at Alcatel-Lucent).

Interdisciplinary work. Partly by design, partly by fortunate coincidences, I have had the chance to work with people from different disciplines. My two most established interdisciplinary efforts are the following:

a. Bionformatics and Computer Networks. I have started collaborating with bionformatics researchers on how to use network analysis techniques on biological problems. For example, I have worked with Stefano Lonardi at UCR [56], B. Andreopoulos then at U. of Toronto [1], and Paul De Ley from the UCR Dept. of Nematology. With Prof. De Ley, we secured a large NSF grant focusing on analyzing nematodes ("Interactive and intelligent searching of biological images by query and network navigation with learning capabilities", $1.1M).

b. Datamining and Computer Networks. My major collaborator here is my older brother Christos Faloutsos, a data-mining expert, with whom I have multiple papers at the cusp of computer networks research and datamining, which are spread among networking and datamining forums (such as ACM KDD [49] and ECML-PKDD [48]). In addition, we have had several NSF and Army grants, partially due to the novel perspective and enabling capabilities of our interdisciplinary approach.

1 Current and future research thrusts

This section presents the existing and ongoing work in each research area mentioned above in more detail. Note that it is only possible to highlight indicative efforts from each area. All the areas listed below are active areas that I intend to continue pursuing in the future, as I briefly discuss at the end of this document.

[1] I collaborated with Pelechrinis in papers, some still under submission, but I did not advise him officially.

1.1 Web security: fighting cyber-crime

Web security is a more recent area of research for my group, which attracted us because of the dire need for protecting both websites and web-surfers. The urgency and its immediate practical impact have increased, since the web seems to have become the vector of choice for virus propagation.

Scope and importance: Every day, 6000 websites are blacklisted (as dangerous) by Google. Websites have become the new battleground of cyber-terrorism and cyber-crime. Web-based crime hurts the reputation and revenue of e-commerce sites, turns websites into distributors of malware, compromises sensitive company and customer data, and threatens national security. An estimated 70-90% of the 75 million active websites are vulnerable, including highly-visible sites. Being professionally hosted is not a solution either: large hosting companies have tens of thousands of compromised sites. Trends suggest that website protection is where PC anti-virus protection was in 1999: what started as an optional luxury, has become an absolute necessity.

Our work so far has focused on: (a) understanding spam, web-based malware, and the web-infrastructure that supports it [53, 42], (b) the detection of parasitic (typosquatting) sites [5], and (c) the detection of phishing sites [46]. Our vision is to become a catalyst for website security research by establishing a comprehensive agenda that includes: developing new approaches, evaluating existing methods, and integrating all these capabilities into a modular and extensible toolkit bringing together the research community and industry.

Practical and commercial impact: My startup has emerged from this area of research: StopTheHacker.com was co-founded by my PhD student A. Banerjee in 2008. The company focuses on protecting websites against hackers by providing preventive and early-diagnosis "health-care" services in a Software as a Service fashion. The company has received $600K in government awards, started offering commercial services since 2009, has recently established key partnerships with webhosting companies, and is preparing for its first institutional investment.

1.2 Network science, complex systems and social networks

Scope and importance: The overarching goal is to understand how large complex systems behave and evolve, which includes methods to model the topology, understand the evolution, identify patterns, and detect anomalies. Another interesting and related question is to understand and model fundamental propagation properties of a network, such as transition phenomena.

Our work addresses the study of networks in a unified way independently of their origin, which includes communication networks, human interactions, biological networks, and online social networks. Our most celebrated work was the discovery of network properties that are described by skewed distributions and powerlaws, focusing on the Internet topology, a work that received the ACM Test of Time award [17]. Since then, we have delved into related issues, including modeling, sampling, and generating realistic Internet-like topologies [30, 29, 27, 28].

More recently, we have expanded the work to include biological networks [56] [1], and online social networks, such as the buyer-seller graph of eBay [6] and video popularity on youtube [10].

Interestingly, this area overlaps with data-mining research, and we have interacted with both communities [49, 48]. One of our most recent thrusts studies fundamental properties of virus propagation in time-varying networks [48] and static networks [9]. We provide the first analytical work that shows that the eigenvalue of an appropriately constructed adjacency matrix of the system determines whether a virus will survive or become extinct.

Our most recent and nascent research effort is the use of network science and graph mining concepts in software engineering. The key challenge is to improve software quality and reduce the cost of maintenance by identifying the most problematic software functions and modules for efficient software maintenance and debugging. The effort proposes a novel approach to this problem based on modeling software as a graph and the use of sophisticated graph-based techniques. Our preliminary results are very promising: using long-lived open-source programs such as Mozilla, we are able to predict bug severity and estimate the software maintenance effort with high accuracy. This effort is in collaboration with Iulian Neamtiu, a new software engineering faculty member at UCR.

1.3 Network security and traffic modeling

Within the large area of network security, we have focused on measuring, modeling, and classifying traffic, and using this knowledge and tools to detect malware and intrusions.

Scope and importance: Monitoring the traffic and detecting unwanted applications is far from trivial. The authors of controversial applications often obfuscate their traffic to make them very hard to detect by using encryption or ever-changing behavior. Thus, we want an approach that has the following properties: (a) it is easy to use

in practice and for a wide range of tasks, and (b) it can operate with limited, partial or erroneous information. Despite the numerous efforts in this direction, the problem is far from solved.

A key novelty of our work is to follow a more fundamental *behavioral* approach, where the detector looks for behavior patterns of the application that are both intrinsic to the application and distinct from other traffic. By identifying intrinsic behaviors, we make it difficult for application writers to disguise their applications without defeating the very purpose of the application.

Our behavioral driven approach has two thrusts, each developed by a different PhD student. Our first approach focuses on the behavior of a single node and characterizes the traffic based on an appropriately constructed graph that captures that nodes interaction with other nodes [41, 40] (more than 500 citations in Google Scholar). In our second approach, we go a step further, and we use the network wide interaction graph, which captures the communication pattern in a network, namely who talks to whom [33] [31] [32] [23]. We develop a plethora of methods that address different challenges and show beyond doubt that there is a wealth of information embedded in a such a graph-based approaches, which goes beyond or complements what previous methods can accomplish. There is still significant ongoing work in exploring the full capabilities of our approaches, and in developing targeted tools with the aid and the direction of our industry partners (Narus, Cisco, Microsoft, Intel), and government and military agencies.

Over the span of 10 years, we have also completed a significant amount of work in the fundamental understanding of the nature of traffic [39], traffic measurement and application classification in general [37, 36], and this includes widely used tools [38] (more than 2ooo downloads within the first three years), and research efforts that have attracted wide interest in the popular press [36].

1.4 Network protocols: robustness, scalability, and efficiency

This broadly-defined area includes a large number of multi-year efforts, and multiple PhD and Master's students. This area includes efforts that attempt to: (a) measure and model how various network protocols work, and (b) design more efficient protocols. Below is a list of several of our efforts grouped for ease of presentation.

1. *Internet interdomain (BGP) routing.* Interdomain routing is the backbone of the Internet, it is what ensures the connectivity between networks belonging to different entities (Automonous Systems). We have studied the behavior of BGP [51, 47], developed methods to improve its robustness [52, 50], and tools to troubleshoot problems [49].

2. *Robust and scalable wireless and ad hoc routing.* We have proposed fundamentally different approaches to routing [12, 13] by proposing the use dynamic addressing, and making a distinction between permanent node identities and transient addresses. We have studied security challenges and proposed secure routing algorithms that can withstand various types of attacks [14, 15, 26, 25]. We have also proposed routing protocols that exploit the latest capabilities and breakthroughs at the physical layer [35, 34].

   We have also developed protocols to improve the performance of wireless infrastructure-based networks [7, 8] and hybrid networks with an infrastructure-based and an ad hoc component [45, 55].

3. *Content distribution and distributed systems.* We have conducted multiple studies on how content can be distributed over a network, and we have studied various aspects including: (a) modeling the effect of misbehaving users and the distribution performance[4], (b) user privacy and anonymity [3] [2] (best paper award and wide publicity in the popular press), and (c) enabling realtime streaming of content in a distributed fashion[54] (147 citations in Google Scholar).

4. *Efficient multicasting and broadcasting.* Regarding group communications, we have studied and developed multicast and broadcast protocols for wireline [16, 22, 11] and wireless networks [44, 43, 24], that attempt to strike the balance in a variety of operational trade-offs. In our earlier work, we have also done theoretical research in distributed algorithms and fundamental bounds on the competitive ratio of algorithms [18, 19, 20, 21].

2 Conclusion and Future Plans

This concludes a brief overview of my work over the last roughly 15 years. Regrettably, several interesting efforts had to be omitted for the sake of brevity.

In the future, I am particularly excited about the first two research areas: web security and network science, while at the same time continuing the work on network security and measurements, and the development of robust routing solutions. I will expand a little more on the first two areas.

My interest in web security is motivated by its importance and urgency, which will only increase in the future. The are two main reasons for this observed trend. First, hackers from all over the world are constantly finding new and ingenuous ways to attack. At the same time, web technology is becoming more complex, interactive, and more accessible, which means that non-technical people can launch their own website in a matter of minutes. My intent is to play a role in making the web a safer place. This will require several things: (a) keeping abreast of the novel techniques hackers use, (b) pulling resources together from diverse technical fields, (c) leveraging the power of community efforts and user participation, and (d) leading efforts to develop standards, best practices and educate users and technologists.

My interest in network science, an admittedly overused term, focuses on understanding the behavior of large systems through the use of sophisticated and novel graph-mining techniques. First, understanding how a complex system functions, and detecting the emergent behavior resonates with the most basic curiosity that made me a scientist in the first place. As I described before, I am interested in many kind of complex systems, including man-made (Internet topology), communication (call and text-message graphs), biological networks (Protein-to-Protein interaction), and human-centric (eBay and youtube). Second, I am interested in exploring the power of graph-based techniques to provide novel insight into the interworkings of complex systems. It is natural to attempt to build on the success we have had with bringing graph-based models and techniques in diverse settings, and leverage the expertise we have developed in this field.

Having said that, I believe that scientists need to be agile: we need to keep our eyes and our mind open and let scientific curiosity and practical needs guide us. At the same time, as anyone who has supervised graduate students knows, the interest and enthusiasm of the students often end up shaping the research agenda, often in wonderfully surprising ways.

References

[1] Bill Andreopoulos, Aijun An, Xiaogang Wang, Michalis Faloutsos, and Michael Schroeder. Clustering by common friends finds locally significant proteins mediating modules. *Bioinformatics*, 23(9):1124–1131, 2007.

[2] A. Banerjee, M. Faloutsos, and L. Bhuyan. Is someone tracking p2p users? In *IFIP NETWORKING (best paper award)*, 2007.

[3] A. Banerjee, M. Faloutsos, and L. Bhuyan. The p2p war: Someone is monitoring your activities. *Computer Networks*, 52(6):1272–1280, 2008.

[4] A. Banerjee, M. Faloutsos, and L. Bhuyan. Profiling podcast-based content distribution. *IEEE Global Internet (Infocom Workshops)*, 2008.

[5] Anirban Banerjee, Dhiman Barman, Michalis Faloutsos, and Laxmi Bhuyan. Cyber-fraud is one typo away. In *IEEE INFOCOM, mini-conference*, 2008.

[6] Yordanos Beyene, Michalis Faloutsos, Duen Horng (Polo) Chau, and Christos Faloutsos. The ebay graph: How do online auction users interact? In *IEEE Global Internet*, 2008.

[7] I. Broustis, K. Papagiannaki, S. V. Krishnamurthy, M. Faloutsos, and V. Mhatre. MDG: Measurement-Driven Guidelines for 802.11 WLAN Design. In *ACM MOBICOM, Montreal, Canada*, 2009.

[8] I. Broustis, K. Papagiannaki, S. V. Krishnamurthy, M. Faloutsos, and V. Mhatre. Measurement-Driven Guidelines for 802.11 WLAN Design. *IEEE/ACM Transactions on Networking*, 18(3), 2010.

[9] Deepayan Chakrabarti, Jure Leskovec, Christos Faloutsos, Samuel Madden, Carlos Guestrin, and Michalis Faloutsos. Information Survival Threshold in Sensor and P2P Networks. *IEEE INFOCOM*, 2007.

[10] Gloria Chatzopoulou, Cheng Sheng, and Michalis Faloutsos. A First Step Towards Understanding Popularity in YouTube. In *Workshop on Network Scienece for Communication Networks (NetSciCom) (In Conjuction with IEEE Infocom 2010)*, 2010.

[11] Jun-Hong Cui, Michalis Faloutsos, Dario Maggiorini, Mario Gerla, and Khaled Boussetta. Measuring and modelling the group membership in the internet. In *ACM SIGCOMM/USENIX Internet Measurement Conference (IMC 2004)*, October, 2003.

[12] J. Eriksson, M. Faloutsos, and S. V. Krishnamurthy. Scalable ad hoc routing: The case for dynamic addressing. In *IEEE INFOCOM, Hong Kong*, 2004.

[13] J. Eriksson, M. Faloutsos, and S. V. Krishnamurthy. DART: Dynamic Address RouTing for Scalable Ad Hoc and Mesh Networks. *Networking, IEEE/ACM Transactions on*, 15(1):119–132, 2007.

[14] J. Eriksson, S. V. Krishnamurthy, and M. Faloutsos. Truelink: A practical countermeasure to the wormhole attack in wireless networks. *IEEE ICNP*, pages 75–84, 2006.

[15] J. Eriksson, S. V. Krishnamurthy, and M. Faloutsos. Routing amid colluding attackers. *IEEE ICNP*, pages 184–193, 2007.

[16] M. Faloutsos, A. Banerjea, and R. Pankaj. QoSMIC: a QoS Multicast Internet protoCol. *ACM SIGCOMM*, Sep 2-4, Vancouver BC 1998.

[17] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the Internet topology. *ACM SIGCOMM*, pages 251–262, Sep 1-3, Cambridge MA, 1999.

[18] M. Faloutsos and M. Molle. Optimal distributed algorithm for minimum spanning trees revisited. *Proceedings of 1995 Principles Of Distributed Computing (PODC)*, 1995.

[19] M. Faloutsos and M. Molle. A linear-time optimal-message distributed algorithm for minimum spanning trees. *Distributed Computing*, 17(2):151–170, August 2004.

[20] M. Faloutsos, R. Pankaj, and K. C. Sevcik. Bounds for the on-line multicast problem in directed graphs. *Proceedings of 4th International Colloquium on Structural Information and Communication Complexity (SIROCCO '97), Monte Verita', Ascona, Switzerland July 24-26*, pages 81–98, 1997.

[21] M. Faloutsos, R. Pankaj, and K. C. Sevcik. The effect of asymmetry on the on-line multicast routing problem. *International Journal on Foundations in Computers Science*, 13(6):889–910, 2002.

[22] A. Fei, J. Cui, M. Gerla, and M. Faloutsos. Aggregated multicast: an approach to reduce multicast state. In *IEEE GLOBECOM Global Internet Symp., San Antonio*, 2001.

[23] Brian Gallagher, Marios Iliofotou, Tina Eliassi-Rad, and Michalis Faloutsos. Homophily in application layer and its usage in traffic classification. In *IEEE INFOCOM*, San Diego, CA, USA, March 2010.

[24] Min Ge, Srikanth V. Krishnamurthy, and Michalis Faloutsos. Application versus network layer multicasting in ad hoc networks: the alma routing protocol. *Ad Hoc Networks*, 4(2):283 – 300, 2006.

[25] V. Gupta, S. V. Krishnamurthy, and M. Faloutsos. Improving the performance of tcp in the presence of interacting udp flows in ad hoc networks. In *IFIP Networking 2004, Athens, Greece*, 2004.

[26] V. Gupta, S.V. Krishnamurthy, and M. Faloutsos. Denial of Service Attacks at the MAC Layer in Wireless Ad Hoc Networks. In *Proc. IEEE MILCOM*, 2002.

[27] Y. He, M. Faloutsos, S.V. Krishnamurthy, and M. Chrobak. Policy-aware topologies for efficient inter-domain routing evaluations. *IEEE INFOCOM mini-conference*, 2008.

[28] Y. He, M. Faloutsos, S.V. Krishnamurthy, and M. Chrobak. Obtaining provably-legitimate internet topologies. *IEEE/ACM Trans. on Networking, to appear*, 2011.

[29] Y. He, G. Siganos, M. Faloutsos, and S.V. Krishnamurthy. *A systematic framework for unearthing the missing links: measurements and impact*, pages 187–200. 2007.

[30] Y. He, G. Siganos, M. Faloutsos, and S.V. Krishnamurthy. Lord of the links: A framework for discovering missing links in the internet topology. *IEEE/ACM Transactions on Networking*, 17(2):391–404, 2009.

[31] Marios Iliofotou, Michalis Faloutsos, and Michael Mitzenmacher. Exploiting dynamicity in graph-based traffic analysis: Techniques and applications. In *ACM CoNEXT*, December 2009.

[32] Marios Iliofotou, Brian Gallagher, Tina Eliassi-Rad, Guowu Xie, and Michalis Faloutsos. Profiling-by-association: A resilient traffic profiling solution for the internet backbone. In *ACM CoNEXT*, Philadelphia, CA, USA, December 2010.

[33] Marios Iliofotou, Prashanth Pappu, Michalis Faloutsos, Michael Mitzenmacher, Sumeet Singh, and George Varghese. Network monitoring using traffic dispersion graphs (tdgs). In *In ACM Internet Measurement Conference (IMC)*, pages 315–320, 2007.

[34] G. Jakllari, S. V. Krishnamurthy, M. Faloutsos, and P. V. Krishnamurthy. On broadcasting with cooperative diversity in multi-hop wireless networks. *IEEE JSAC special issue on Cooperative Communications and Networking*, 25(2):484–496, Feb 2007.

[35] G. Jakllari, S. V. Krishnamurthy, M. Faloutsos, P. V. Krishnamurthy, and O. Ercetin. A framework for distributed spatio-temporal communications in mobile ad hoc networks. In *IEEE INFOCOM 2006, Barcelona, Spain*, 2006.

[36] T. Karagiannis, A. Broido, N. Brownlee, kc claffy, and M. Faloutsos. Is p2p dying or just hiding? In *IEEE Globecom 2004 - Global Internet and Next Generation Networks*, Dallas, November, 2004.

[37] T. Karagiannis, A. Broido, M. Faloutsos, and kc claffy. Transport layer identification of p2p traffic. In *ACM SIGCOMM/USENIX Internet Measurement Conference (IMC 2004)*, Italy, October, 2004.

[38] T. Karagiannis, M. Molle, and M. Faloutsos. A User-Friendly Self-Similarity Analysis Tool. *ACM Computer Communication Review*, 33(3):81–93, 2004.

[39] T. Karagiannis, M. Molle, M. Faloutsos, and A. Broido. A Nonstationary Poisson View of Internet Traffic. In *IEEE INFOCOM*, Hong Kong, 2004.

[40] T. Karagiannis, D. Papagiannaki, N. Taft, and M. Faloutsos. Profiling the end host. *Passive and Active Measurement Conference (PAM)*, 2007.

[41] T. Karagiannis, K. Papagiannaki, and M. Faloutsos. BLINC: Multi-level Traffic Classification in the Dark. In *ACM SIGCOMM*, August 2005.

[42] M. Kokkodis and M. Faloutsos. Spamming botnets: Are we losing the war? *6th Conference on Email and AntiSpam CEAS 2009*, 2009.

[43] L.K. Law, S.V. Krishnamurthy, and M. Faloutsos. A novel adaptive protocol for lightweight efficient multicasting in ad hoc networks. *Computer Networks Journal to appear*, 51(3):823–834, Feb 2007.

[44] L.K. Law, S.V. Krishnamurthy, and M. Faloutsos. Understanding and exploiting the tradeoffs between broadcasting and multicasting in mobile ad hoc networks. *IEEE Trans. on Mobile Computing to appear*, 6(3):264–279, March 2007.

[45] L.K. Law, K. Pelechrinis, S.V. Krishnamurthy, and M. Faloutsos. Downlink capacity of hybrid cellular ad hoc networks. *IEEE/ACM Transactions on Networking*, 18(1):243–256, 2010.

[46] Anh Le, Athina Markopoulou, and Michalis Faloutsos. Phishdef: Url names say it all. *IEEE INFOCOM*, mini conference, 2010.

[47] C. Palmer, G. Siganos, M. Faloutsos, C. Faloutsos, and P. Gibbons. The connectivity and fault-tolerance of the Internet topology. *Workshop on Network-Related Data Management (NRDM 2001), In cooperation with ACM SIGMOD/PODS, Santa Barbara*, 2001.

[48] B. Aditya Prakash, Hanghang Tong, Nicholas Valler, Michalis Faloutsos, and Christos Faloutsos. Virus propagation on time-varying networks: Theory and immunization algorithms. In *ECML/PKDD (3)*, pages 99–114, 2010.

[49] B. Aditya Prakash, Nicholas Valler, David Andersen, Michalis Faloutsos, and Christos Faloutsos. Bgp-lens: patterns and anomalies in internet routing updates. In *ACM KDD*, pages 1315–1324, 2009.

[50] G. Siganos and M. Faloutsos. Neighborhood Watch for Internet Routing: Can we improve the Robustness of Internet Routing Today? *IEEE INFOCOM*, 2007.

[51] G. Siganos and M. Faloutsos. BGP Routing Properties at a Large Time Scale. In *IEEE GLOBECOM, Global Internet Symposium*, TAIPEI,2002.

[52] Georgos Siganos and Michalis Faloutsos. Analyzing BGP policies: Methodology and tool. *IEEE INFOCOM*, 2004.

[53] M. Faloutsos TK Huang, Nicholas Valler. Characterizing the scam hosting infrastructure. In *IEEE GLOBECOM*, 2010.

[54] A. Vlavianos, M. Iliofotou, and M. Faloutsos. Bitos: Enhancing bittorrent for supporting streaming applications. *IEEE Global Internet, colocated with IEEE INFOCOM*, 2006.

[55] Serdar Vural, Lap Kong Law, Srikanth V. Krishnamurthy, and Michalis Faloutsos. On the uplink capacity of hybrid cellular ad hoc networks. In *IEEE SECON, June 2010, Boston, MA, USA (best paper award)*, 2010.

[56] Q. Yang, G. Siganos, M. Faloutsos, and S. Lonardi. Evolution versus Intelligent Design: Comparing the Topology of Protein-Protein Interaction Networks to the Internet. In *LSS Computational Systems Bioinformatics Conference (CSB'06) Stanford, CA,*, pages 299–310, August 2006.