BGP, Not As Easy As 1-2-3

Ashley Flavel

Thesis submitted for the degree of

Doctor of Philosophy

in

Applied Mathematics

at

The University of Adelaide

Faculty of Engineering, Computer and Mathematical Sciences

October 6, 2009

Signed Statement

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

SIGNED: . . . . . DATE: . . . . . .

Acknowledgements

I would firstly like to acknowledge the support of my primary supervisor, Associate Professor Matthew Roughan. It was he who first inspired me to pursue a researching career. In addition, his financial and social assistance helped gain me an internship at AT&T Research that reinvigorated my desire to work on real-world problems with real-world solutions.

I would also like to acknowledge the support of my co-supervisor, Professor Nigel Bean. His ability to interpret my sometimes incoherent description of a problem into one that was simple and easy-to-understand was a major factor in the success of this thesis.

Dr. Olaf Maennel, as fellow co-supervisor, provided an alternative perspective to inter-domain routing to my primary supervisor A/Prof Roughan. His pragmatic approach to inter-domain routing and in-depth knowledge of its quirks and corner cases was an excellent source to verify the sanity of my ideas. All three of my supervisors provided contrasting perspectives which although at times was frustrating allowed the development of ideas that were not only theoretically sound but pragmatic. I would like to thank them all for their friendship and support over the last few years.

Dr. Aman Shaikh of AT&T Research Labs was a major contributor to the ideas in this thesis. The four months I spent at AT&T Research and the subsequent year-long collaboration resulted in the best work of my thesis. Aman also assisted with my transport to and from the laboratories in which time many of the most important research breakthroughs came (as well as many discussions on cricket). AT&T also deserve acknowledgement for providing me access to their

commercially sensitive network data.

This thesis was reviewed by three of my most admired networking researchers:- Dr. Timothy Griffin of The University of Cambridge, Professor Jennifer Rexford of Princeton University and Dr. Steve Uhlig of Technische Universität Berlin. I am sincerely privileged and greatly appreciate their time and useful comments.

For financial assistance I am very grateful to the Australian Research Council's Communications Research Network (ACoRN). ACoRN provided me with substantial financial assistance for my research visit to AT&T. I specifically would like to thank ACoRN's Adelaide University representative Belinda Chiera. In addition, I would like to acknowledge the financial support of the Australian Research Council through grant DP0557066.

Dr. Flo Rice, although not related to my technical research, was a pivotal figure in its development. She provided me with accommodation for the time I spent at AT&T and became a good friend.

The staff and students at the Teletraffic Research Centre at the University of Adelaide provided me with an excellent research atmosphere. I would specifically like to thank Dr. Jeremy McMahon for frequent useful discussions.

I would like to thank Maxine and Mark Wong See for proof reading several sections of this thesis. A special acknowledgement is reserved for Carrie Kelly who devoted a significant amount of time and energy to proof read the entire thesis.

Lastly, I would like to thank my family for their support throughout my time as a student. Without them my chance to reach this point would not be possible. They have enabled me to reach a level of education that will give me incredible future opportunities.

Dedication

I dedicate this thesis to my family. My desire to succeed comes from them, and without their love and support this thesis would not have been possible.

Contents

List of Figures

List of Tables

Abstract

The Internet is literally an "Inter-Network", that is, a network of networks. Networks can be entities including Internet Service Providers (ISPs), universities and commercial enterprises. Every network or Autonomous System (AS) has individual requirements, restrictions and capabilities to transit data traffic. No central controlling body determines how ASes connect — instead contractual agreements are established between AS pairs to govern their relationship. It is not feasible for all ASes to be physically connected to all others. Consequently, some ASes provide transit between other ASes. Such a service usually results in remuneration from one or both ASes.

Unlike centrally administered networks where all nodes in the network make generic, predictable decisions, each AS has the ability to select its best route based on its own proprietary commercial agreements. Such agreements are converted to a technical policy implemented in the Border Gateway Protocol (BGP). The ability to implement policies ensures the commercial viability of the Internet, but also makes the prediction of routes difficult and even more worrisome, conflicting policies can cause undesirable BGP states where no single AS has sufficient knowledge to understand what is happening [43].

Designing new clean-slate routing protocols is one approach to improving the predictability and reliability of the Internet. However, due to the Internet's distributed political and administrative control, significant collaboration is required to implement a new routing protocol — especially when no new protocol currently proposed has sufficiently superior flexibility, scalability or robustness. The difficulty in implementing new and improved protocols is evident in the deploy-

ment of IPv6 [23]. Although the IPv6 standard has been defined for over a decade and offers a larger address space, better security and embedded quality of service in comparison to traditional IPv4, its deployment is limited to 1200 of over 30000 ASes in the Internet [66]. Hence, it is crucial practical solutions to current problems are evolved in addition to developing clean-slate techniques. Consequently, our approach is pragmatic — designing tangible solutions to practical problems that can be implemented immediately.

In this thesis we examine and combine eBGP, iBGP, OSPF, Netflow and router configuration data to discover important aspects of routing. It is this investigation that instigated the development of a model of iBGP. iBGP is the version of BGP implemented *within* ASes to propagate routes between internal routers. It exists on a logical topology, however it interacts with the physical topology. It is this interaction which can cause persistent oscillation [49] — a system state where routers alter their decision ad infinitum. Detecting configurations which can cause this oscillation is NP-hard [49]. However, our model of iBGP introduced in this thesis benefits from the 'designed' structure of the iBGP topology to restrict the search space dramatically to one that is computationally feasible.

iBGP data — which is collected to analyze router decisions — is often only collected on a subset of routers due to its massive storage requirements. In addition there is a substantial amount of correlation between router decisions. Our model of iBGP discovers the dependencies between router decisions and can consequently predict the decisions of those routers for which no measurements are available. It does not rely on any assumption of operator configuration, and subsequently is able to be applied in any network scenario — not just the one originally configured. It is this feature, together with the model's ability to use any available measurement data that makes our technique ideal for network measurement and management applications. We found our model is efficient and accurate on the network of a large Tier-2 AS, where all but seven of over 12.7 million decisions were consistent with observed data. Further we were able to predict the decision of 85% of routers where observed data was unavailable.

During our analysis, we also identified several minor configuration errors on operational routers when we predicted the "correct" outcome.

The internal state of a network can be influenced by neighboring ASes. Peering agreements are closely guarded due to their commercially sensitivity. They are implemented in BGP in the form of policies and are difficult to infer with publicly available data sources. We examined the peering policies of over 100 ASes from the perspective of a large Tier-2 AS, finding 22% differ from the canonical peering policy outlined in many peering agreements. When a policy differes from the canonical peering policy, it may result in sub-optimal routing within the Tier-2 AS. We used our model of iBGP to firstly predict the decisions of all routers under the current peering policy, before determining the changes that would have occurred under a canonical peering policy. This analysis not only provided a metric for the routing impact of a peers' non-canonical policy, but subsequently used in combination with traffic data allowed us to determine the influence of the peer on traffic flows. Our techniques described allow an AS to fully quantify the impact of a non-canonical peering policy and adapt business arrangements appropriately.

Throughout our analysis of BGP data, we noticed several inconsistencies in the data. Although the results in the above work were insensitive to such inconsistencies, other applications requiring accurate, fine time-scale analysis of the routing state are much more sensitive. Consequently, we undertake a self-consistency check on the BGP data and examine the possible causes of such inconsistencies. We also present a mechanism to 'clean' the data to minimize the effects of any inconsistency.