On Weather and Internet traffic demand

Juan Camilo Cardona[1,3], Rade Stanojevic[2], and Rubén Cuevas[1]

[1] Institute IMDEA Networks
[2] Telefonica Research
[3] UC3M

Abstract. The weather is known to have a major impact on demand of utilities such as electricity or gas. Given that the Internet usage is strongly tied with human activity, one could guess the existence of similar correlation between its traffic demand and weather conditions. In this paper, we empirically quantify such effects. We find that the influence of precipitation depends on both time of the day as well as time of the year, and is maximal in the late afternoon over summer months.

1 Introduction

The analysis and forecasting of the Internet traffic is a well studied topic with a large number of applications [5]. Such studies have used statistical tools to capture the dominant characteristics of the dynamics, without explicitly modeling the dependence with external factors (e.g. social events, weather) that are typically accounted as noise. While it has been known that these factors have a significant impact on the demand of utilities [4] or TV ratings [6], their relationship with the Internet traffic demand is not well understood. In this paper we empirically study the relationship between the Internet traffic demand and one of the factors that plays a significant role in traffic variability: weather.

The interaction between the weather conditions and the traffic demand happens on several timescales. Short term weather events, like precipitations, have a direct effect on the traffic demand. Longer term effects, reflected through seasonal changes in temperature and daylight duration, have a slower influence on the Internet traffic. Here we study the short-term correlations. For the long-term correlation between the traffic and weather and a deeper analysis of the short-term effects we refer the interested reader to our technical report [3].

2 Datasets description

As indicator of the Internet traffic demand in a particular area we use the traffic data from three Internet eXchange Points (IXP): the Slovak-IX, FICIX and INEX. We obtained 5-minute granular traffic from each IXP by storing and processing their publicly available mrtg images. Our Internet traffic dataset includes 8 months of data from INEX and 18 months of data from Slovak-IX and FICIX. Different from large IXPs [1], the traffic from these IXPs is highly local and thus appropriate for our analysis.

To obtain weather data, we use the data provided by the Weather Underground, an easily accessible database available at `http://www.wunderground.`

`com/`. The `wunderground.com` publishes a considerable number of weather parameters with a granularity of 30 minutes. For the sake of this paper, we fetched from this website the precipitation data for the cities where each IXP is located over the period that covers our traffic data.

3 Short term correlations

The data described in Section 2 allows us to notice changes that happen on the traffic of the three localities over short-time scales and compare them to the weather conditions. For that purpose we split the time into 2-hour time-slots. We denote by $u(t)$ the total traffic transiting through the IXP. In order to remove the seasonal effects we normalize $u(t)$ with the average traffic over a two week period centered at t:

$$\bar{u}(t) = \frac{u(t)}{average(u(t-84), \ldots, u(t+84))}.$$

Thus the normalized traffic $\bar{u}(t)$ measures the variability of the traffic on the short-term timescale, without the impact of long-term seasonality observable in some regions.

For each 2-hour time-slot t there are 4 or more weather records in our dataset. We set a binary variable $wet(t)$ to be 1 if any of the weather records reports precipitation (e.g. snow, shower, rain, storm) otherwise we set $wet(t) = 0$. This binary variable helps us simplify the exposition of the results. Our goal is to examine whether precipitation impacts the traffic, and quantify its effect. To that end, we split the day in twelve 2-hour intervals, and calculate average normalized traffic with and without precipitation for each of the twelve intervals:

$$A(i) = \frac{\sum\limits_{mod(s,12)=i} \bar{u}(s)wet(s)}{\sum\limits_{mod(s,12)=i} wet(s)} \quad B(i) = \frac{\sum\limits_{mod(s,12)=i} \bar{u}(s)(1-wet(s))}{\sum\limits_{modd(s,12)=i} (1-wet(s))} \quad i = 0..11$$

thus for the twelve time intervals $0h-2h, 2h-4h, \ldots, 22h-24h$, $A(i)$ and $B(i)$ represent the average normalized load in the interval $[2ih, (2i+2)h]$ with and without precipitation, respectively.

In Figure 1 we depict the values of $A(i)$ and $B(i)$ for the three IXPs. To determine whether the difference between $A(i)$ and $B(i)$ is statistically significant to claim that the means of the samples with and without precipitation are different, we use Welch's t-test [8], which is well-suited for this case as the number of samples for each random variable is different and relatively large. Figure 1 also includes the interval outside of which Welch's t-test rejects the null-hypothesis for a significance level of 0.05. Thus from early afternoon to early evening, with 95% of confidence we can affirm for all IXPs that the mean normalized traffic is larger in timeslots with precipitation than in timeslots without precipitation. For the other periods of the day, the difference between the means is not statistically significant to support that precipitation impacts the traffic.

Finally, we observe that the impact of precipitation is not uniform across the year. Namely, in Figure 2 we depict the relative increment of precipitation

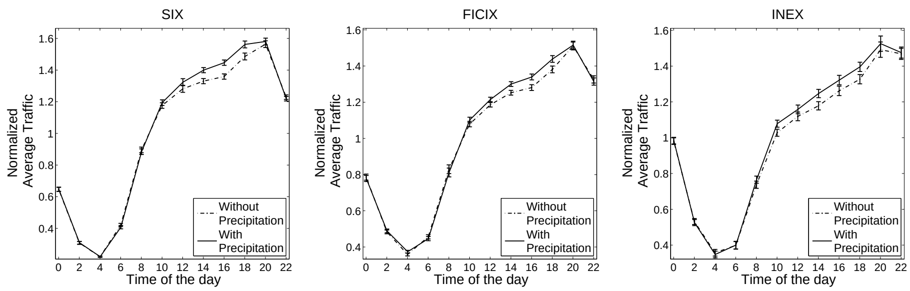

Fig. 1. Normalized daily demand of SIX, FICIX and INEX, with and without precipitation.

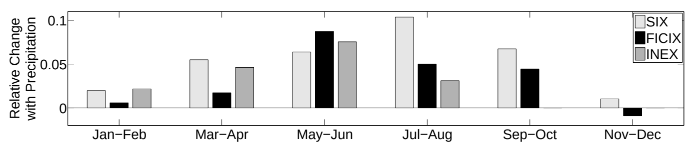

Fig. 2. The relative change with precipitation during the $16h - 18h$ slot over the year.

during the $16h - 18h$ interval for the 6 two-month periods and observe that the impact of precipitation is most pronounced in the summer months, while it is insignificant over the winter.

4 Conclusions

In this paper we examined the dependence between the Internet traffic and the weather in short scales. While for other types of utilities the impact of external factors has been studied in depth, our understanding on such relationship in the Internet is very immature. The phenomena observed here is a step towards filling that knowledge gap and affirms our conjecture that measurable external factors are strongly related with the variability of the Internet traffic. Our work complements other studies that analyze the impact of natural events on the Internet [7, 2]. We refer the reader to [3] for a more extensive analysis of the impact of weather in Internet traffic over short and long scales.

References

1. B. Ager, et al. "Anatomy of a large european IXP." In Proc. of ACM SIGCOMM 2012.
2. Z.S. Bischof, J. S. Otto and F. E. Bustamante. "Distributed Systems and Natural Disasters." Proc. ACM SWID (2011).
3. J. C. Cardona, R. Stanojevic, R. Cuevas. "On Weather and Internet Traffic Demand". T. Report, `https://svnext.networks.imdea.org/repos/WeatherAndInternet/TechReport.pdf`, Sept. 2012.
4. E. Feinberg and D. Genethliou. "Load forecasting". Applied mathematics for restructured electric power systems, 2005.
5. K. Papagiannaki, N. Taft, Z.L. Zhang, C. Diot. "Long-Term Forecasting of Internet Backbone Traffic: Observations and Initial Models". In Proc. of IEEE INFOCOM 2003
6. K. Roe and H. Vandebosch (1996). "Weather to view or not: That is the question". European Journal of Communication, 11(2), 201-216.
7. A. Schulman, N. Spring. "Pingin' in the Rain". In Proc. of ACM IMC 2011.
8. B. L. Welch, "The generalization of Student's problem when several different population variances are involved". Biometrika 34 (1-2): 28-35. doi:10.1093/biomet/34.1-2.28. MR19277. 1947