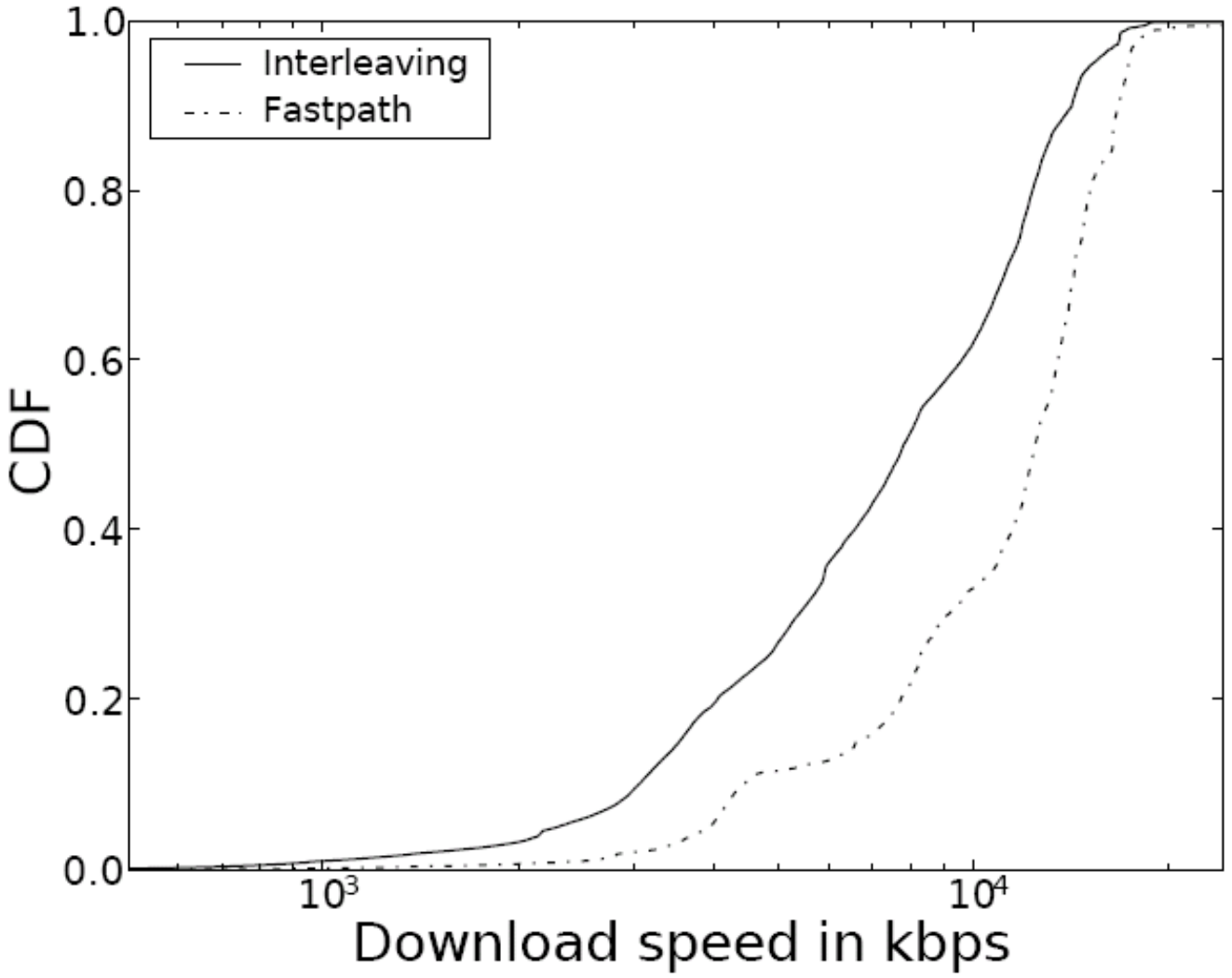
Which factors affect user performance?

- Modem Configuration
- City
- Choice of access ISP
- Service Level Agreement
- DSLAM
- Time of Day
- Day of Week

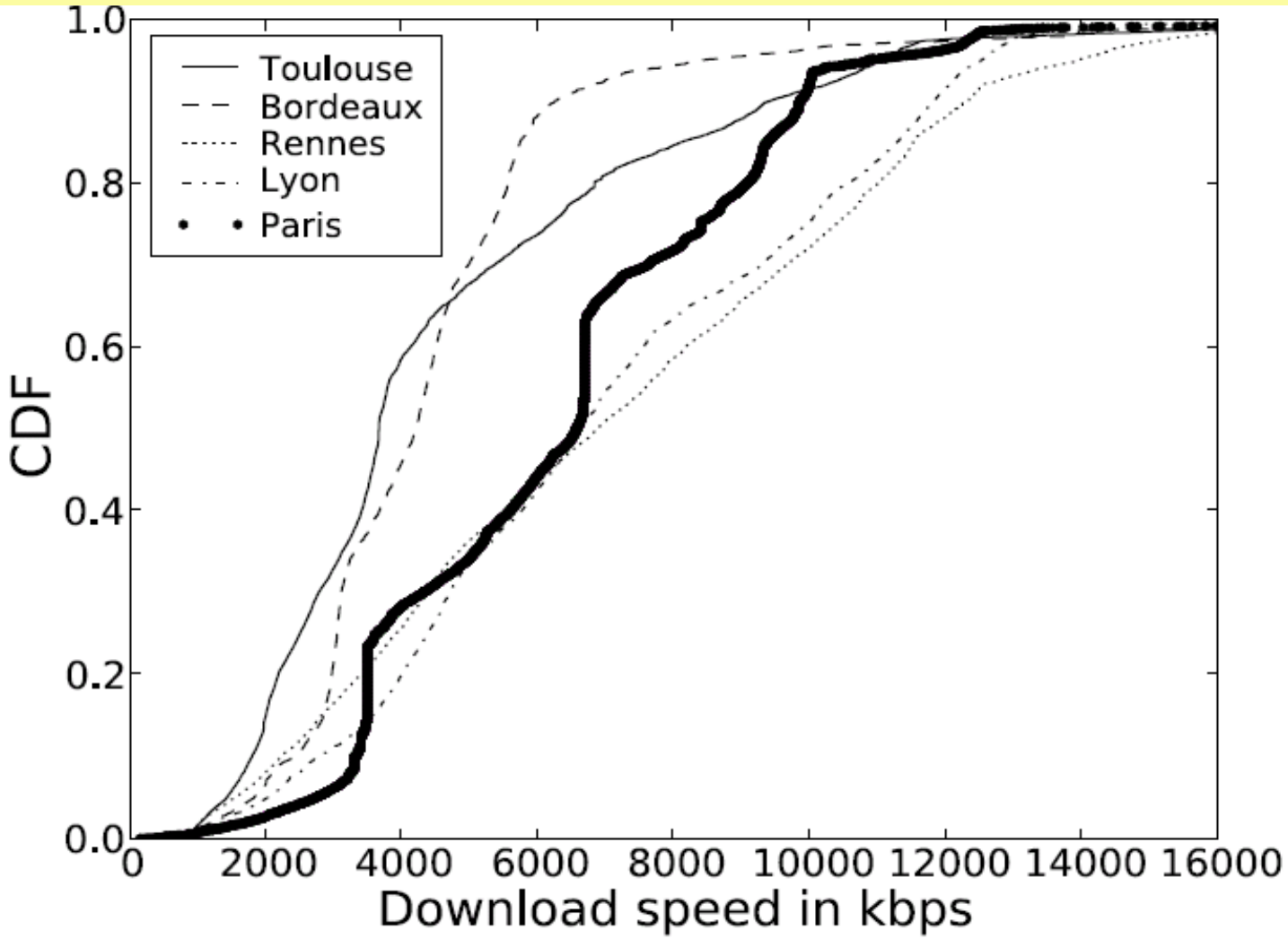
Effects of Modem Configuration

Users who have configured fastpath achieve better download speeds.



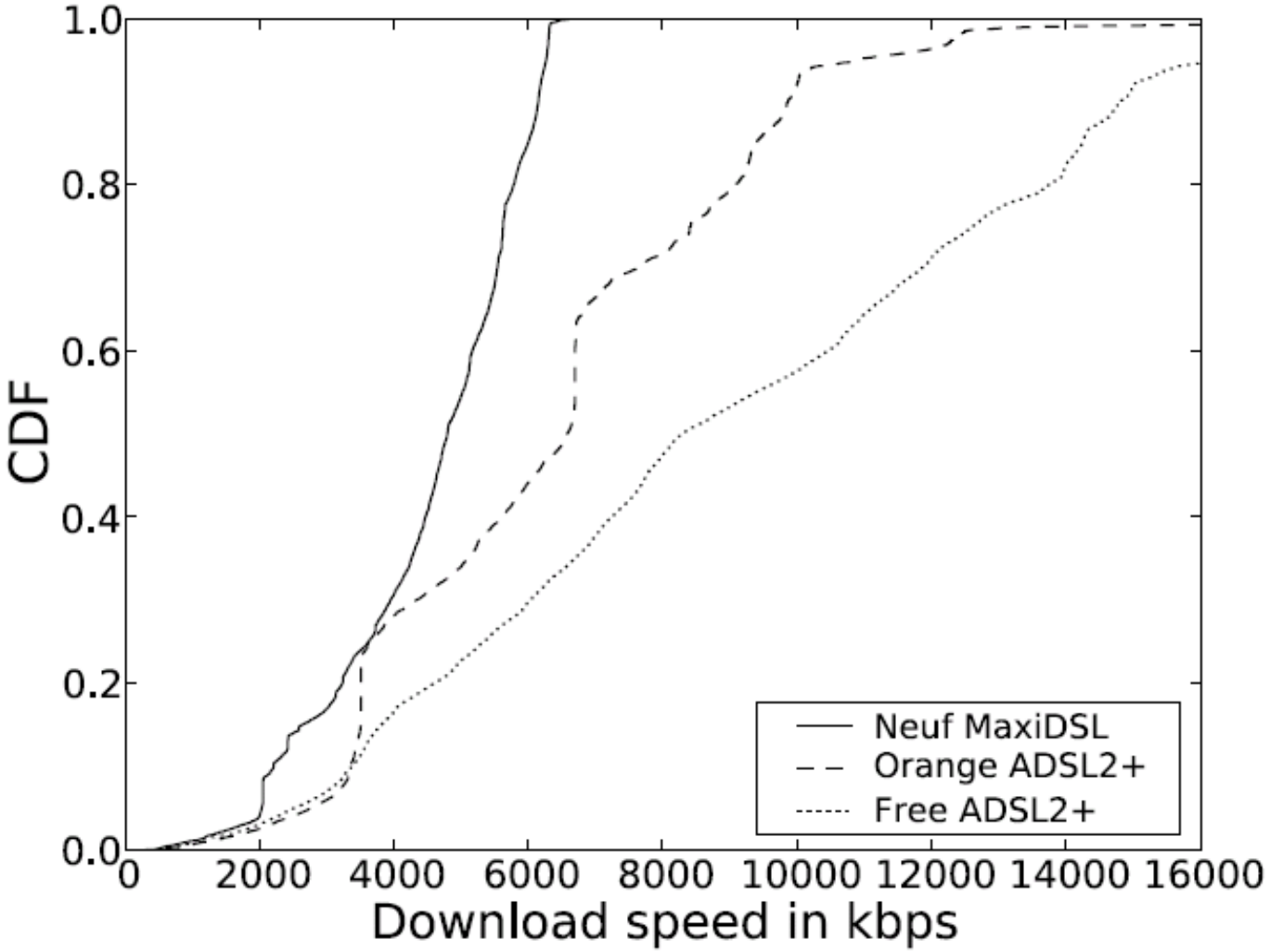
Effect of City

A user's city affects download time significantly.
These differences do not correlate with differences in round-trip time.



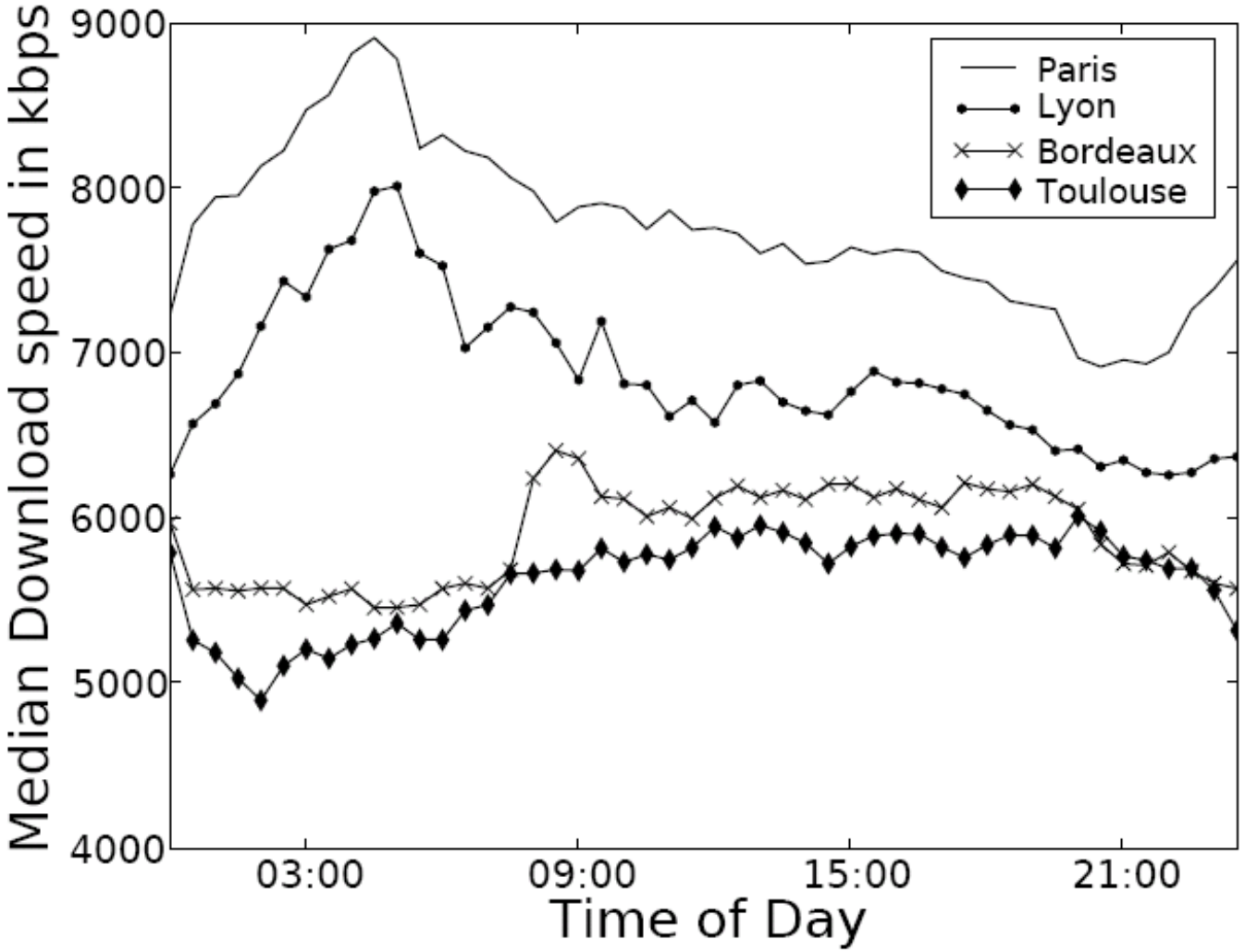
Effect of ISP

Users in the same city and comparable SLAs can experience widely varying performance, for different ISPs.



Effect of Time of Day

Download speeds vary by time of day.



Ranking of Features

- **Question:** Which of the above features are most important in predicting user performance.
- **Approach:** Use ensemble learning to train a predictor of user performance, using these features as input.
- **Output:** predictor and ranking of features

Ensemble Learning: RuleFit

- RuleFit (ensemble learning)
  - $F(x) = a_0 + \sum_{m=1}^{M} a_m f_m(x)$
  - $F(x)$ is the prediction result (label score)
  - $f_m(x)$ are base learners (usually simple rules)
  - a_m are linear coefficients
- Example

	$F(x)$	a_m	$f_m(x)$
Rule 1	**0.080**	0.080	Geodesic distance > 63 AND AS in (1901, 1453, …)
Rule 2	**+ 0**	0.257	Port status: no SMTP service listening

Ranking of Feature Importance

Rank	Upload	Download
1	RTT (100%)	RTT (100%)
2	City (28.5%)	DSLAM (36.6%)
3	DSLAM (25.2%)	Advertised Rate (SLA) (33.7%)
4	Advertised Rate (SLA)(18.3%)	City (32.5%)
5	Fastpath (9.7%)	Time of Day (2.4%)
5	TIme of Day (2.1%)	Fastpath (0.4%)
6	Day of Week (0%)	Day of Week (0%)

- RTT is most important predictor
- DSLAM, City, SLA are also important
- Temporal features are considerably less important

How does performance correlate across time?

- **Question:** When groups of users experience performance fluctuations, what do they share in common?

- **Approach:** Apply cross-correlation and pairwise hierarchical clustering to group users.

Results: Correlated Members

Users from the same ISP experience similar fluctuations, even if they are in different cities.

Member 1	Member 2	Correlation coefficient
Lyon, Free ADSL2+	Paris, Free ADSL2+	0.76
Toulouse, Free 10M Unbundled	Paris, Free ADSL2+	0.56
Lyon, Free ADSL2+	Bordeaux, Free ADSL2+	0.56
Bordeaux, Free ADSL2+	Paris, Free ADSL2+	0.51
Lyon, Free ADSL2+	Toulouse, Free 10M Unbundled	0.50
Paris Orange ADSL2+	Paris Orange ADSL Max	0.47
Lyon Orange ADSL Max	Paris Orange ADSL Max	0.46
Lyon Orange ADSL Max	Paris Orange ADSL2+	0.42
Bordeaux, Free ADSL2+	Toulouse, Free 10M Unbundled	0.42
Lyon Orange ADSL2+	Paris Orange ADSL Max	0.40

Conclusion

- So far, mostly expected results
  - ISPs often do not meet their SLAs
  - SLA is a good indicator of performance
  - ISP is a good predictor of performance fluctuation

- Next steps
  - Deployment: gather more detailed measurements
  - Application: Can correlation help identify root cause?