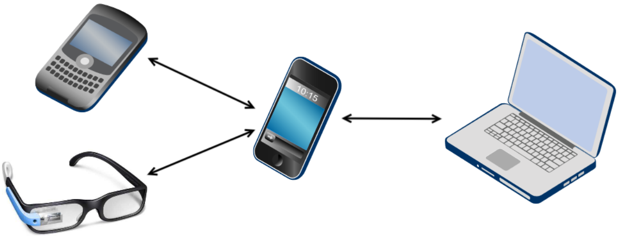
Detecting spatio-temporal clusters in Reality Mining datasets

Matthew Orlinski
Fraunhofer IAIS

March 31, 2014

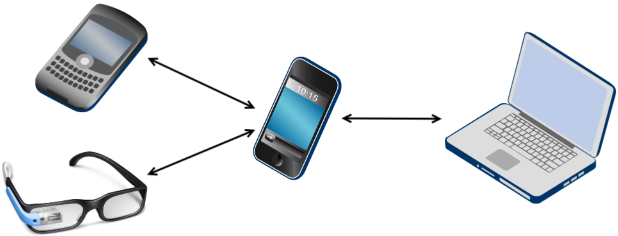
OppNets

- Opportunistic Networks (OppNets) are made up of mobile devices communicating directly with each other.



OppNets

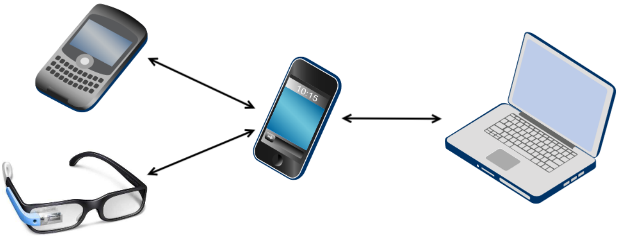
- Opportunistic Networks (OppNets) are made up of mobile devices communicating directly with each other.



- They can be used for:
  - Pervasive sensing applications.

OppNets

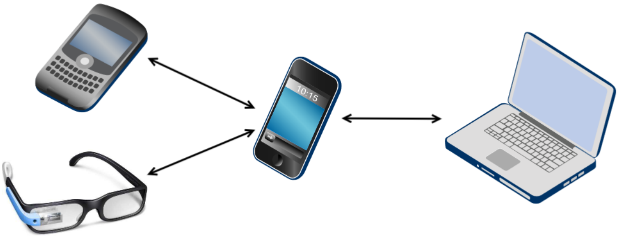
- Opportunistic Networks (OppNets) are made up of mobile devices communicating directly with each other.



- They can be used for:
  - Pervasive sensing applications.
  - Sharing resources.

OppNets

- Opportunistic Networks (OppNets) are made up of mobile devices communicating directly with each other.



- They can be used for:
  - Pervasive sensing applications.
  - Sharing resources.
  - Distributing locally relevant information.

Challenges for OppNets

- Connections are intermittent, making multi-hop communication between participants difficult.

Challenges for OppNets

- Connections are intermittent, making multi-hop communication between participants difficult.

- We can't use many of the protocols developed for MANETS because of the high level of device mobility.
- Instead we use Delay or Disruption Tolerent (DTN) protocols.

What can we do?

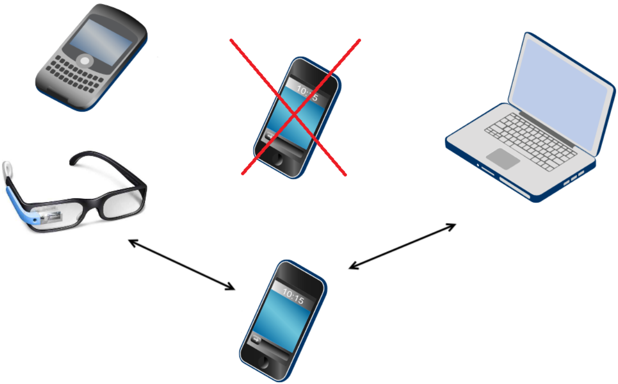
- Analysing how humans interact may help to improve the performance of OppNets.

What can we do?

- Analysing how humans interact may help to improve the performance of OppNets.
  1. Outline the energy requirements for future Peer Aware Communications (PAC) standards such as IEEE 802.15.8 and Wi-Fi direct.

What can we do?

- Analysing how humans interact may help to improve the performance of OppNets.
  1. Outline the energy requirements for future Peer Aware Communications (PAC) standards such as IEEE 802.15.8 and Wi-Fi direct.
  2. **Improve data routing protocols by taking advantage of human mobility to forward packets.**



Reality mining data

	Infocom5	Infocom6	Cambridge	Reality
Environment	Conference		Campus	
Duration (days)	3	3	12	246
Number of devices	41	78	36	97
Device type	iMote	iMote	iMote	Phone
Number of encounters	22459	128979	10641	102594
Daily encounter probability	0.78	0.73	0.24	0.01
Granularity (seconds)	120	120	600	300
Geographic location	No	No	No	Cell ID

Table: Comparison of some reality mining datasets. The daily encounter probability is the probability that an encounter with a particular other device will take place on any given day. Granularity is the time between peer discovery scans.

Reality mining data

	Infocom5	**Infocom6**	**Cambridge**	**Reality**
Environment	Conference		Campus	
Duration (days)	3	3	12	246
Number of devices	41	78	36	97
Device type	iMote	iMote	iMote	Phone
Number of encounters	22459	128979	10641	102594
Daily encounter probability	0.78	0.73	0.24	0.01
Granularity (seconds)	120	120	600	300
Geographic location	No	No	No	Cell ID

Table: Comparison of some reality mining datasets. The daily encounter probability is the probability that an encounter with a particular other device will take place on any given day. Granularity is the time between peer discovery scans.

Visually analysing the encounters with a graph

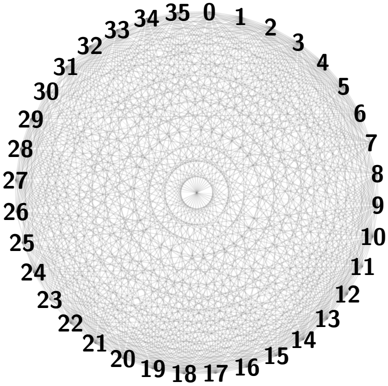
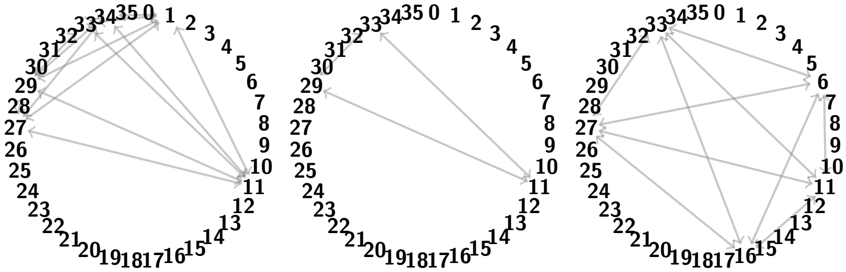


Figure: All of the encounters in the Cambridge dataset.

Discrete time frames



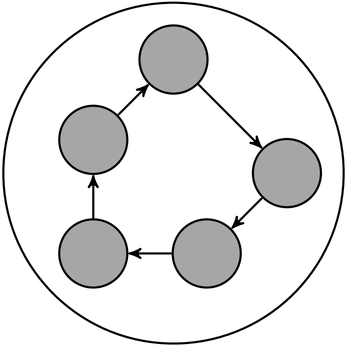
(a) 4^{th} hour (b) 5^{th} hour (c) 6^{th} hour

Figure: The encounters present in the 4^{th}, 5^{th}, and 6^{th} hourly time frames of the Cambridge dataset.

We experimented with discrete time frames of 1, 6, and 12 hours.

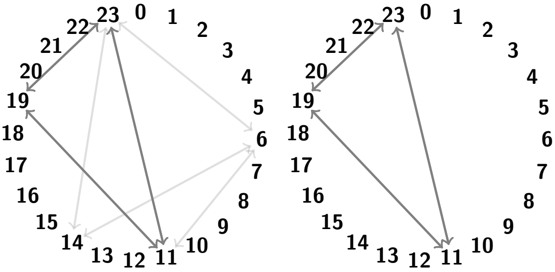
Strongly connected subgraphs

- Strongly connected subgraphs are section of a directed graph where there is a path from every vertex to every other vertex.



Multi-frame spatio-temporal clusters (MEBS)

- **MEBS are strongly connected subgraphs which are present across multiple consecutive time frames**.
- Stable structures which could form a backbone for the network.



(a) x^{th} hour (b) $x + 1^{th}$ hour

Figure: A MEBS cluster formed over 2 time frames.

Single-frame Expectation-Based Spatio-temporal clusters (SEBS)

- **SEBS are strongly connected subgraphs in a single frame where the value of some metric is higher than expected for every node**.
- Used to show when people come together for a short amount of time.
- The expected values are called the baselines.

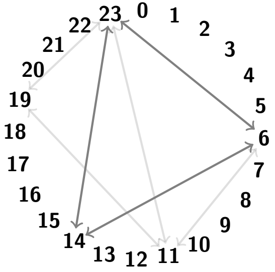


Figure: SEBS example in the x^{th} hour

Baselines for devices

$$\boxed{b_i^{t_n}} = \frac{m_i^{t_{(n-1)}} + m_i^{t_{(n-2)}} + \cdots + m_i^{t_{(n-w)}}}{w}$$

- The baseline for a single device i during the current time frame $t_{(n)}$.

Baselines for devices

$$b_i^{t_n} = \frac{\boxed{m_i^{t_{(n-1)}}} + \boxed{m_i^{t_{(n-2)}}} + \cdots + \boxed{m_i^{t_{(n-w)}}}}{w}$$

- Baselines are calculated using the average cumulative connection time with other devices in the previous w hours.
- We looked at the cumulative connection time because single connections between devices are unstable (the Parking Lot Problem).
- Only storing the average cumulative connection time for each time frame also cuts down on the amount of storage and processing needed for the algorithm.

Baselines for devices

$$b_i^{t_n} = \frac{m_i^{t_{(n-1)}} + m_i^{t_{(n-2)}} + \cdots + m_i^{t_{(n-w)}}}{w}$$

- We experimented with w values of 2, 4, 12, 18, and 24.

A short cluster analysis...

Size of SEBS clusters

- Mean size is around 5-6 devices.
- With frame sizes of 12 hours there are large clusters detected at the start of Infocom6, and there are no clusters detected at all in Infocom5.
- This is due to the **premature clustering problem** which is caused by high metric values and mixing of participants at the start of the experiment, with ever decreasing participation thereafter.

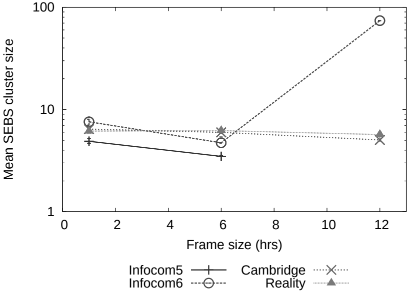


Figure: Mean size of SEBS clusters

Size of MEBS clusters

- MEBS clusters at conferences tend to be smaller than in campus experiments.

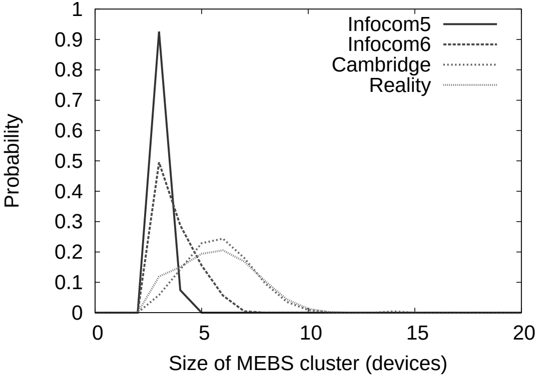
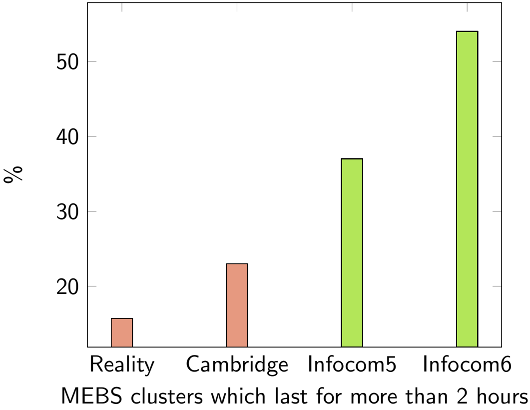


Figure: Size of MEBS clusters spanning 2 time frames when using 1 hour long frames.

Duration of MEBS clusters

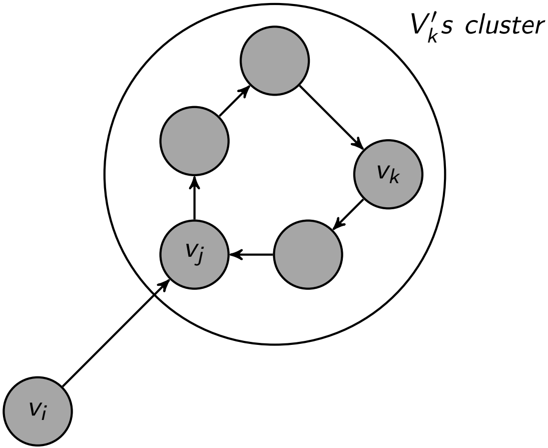
- MEBS clusters which last longer than 2 hours are rare in each of the datasets.
- MEBS clusters at conferences tend to last longer than in campus experiments.



Short analyses summary

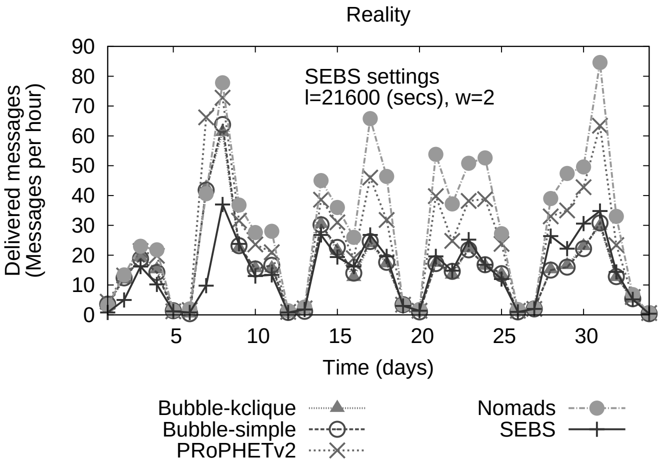
- We've outlined a simple method to detect spatio-temporal clusters.
- Can SEBS clusters meet our requirements? (next slide)

Cluster based data delivery



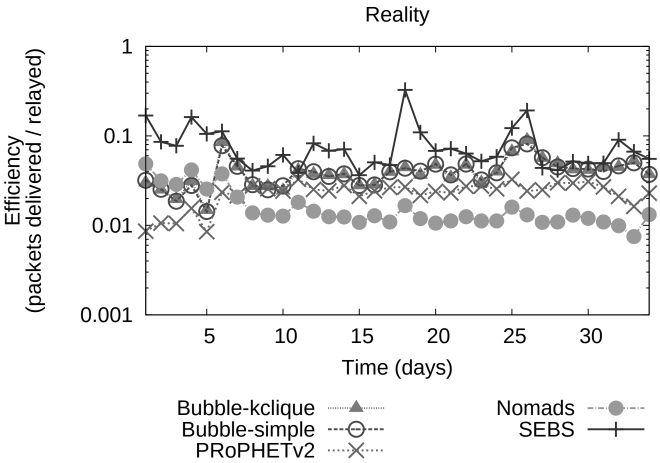
End-to-end data delivery based on SEBS clusters

- Low data delivery probability because SEBS clusters tend to be too small.



End-to-end data delivery based on SEBS clusters

- SEBS clusters are more efficient then Bubble's because SEBS clusters are **non-monotonic**.



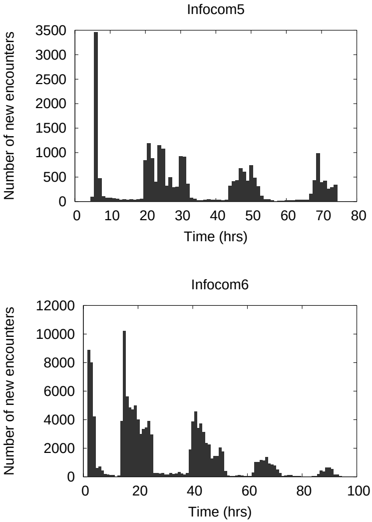
Data delivery overall

	Data delivery probability (TTL 1 hour)	**Overheads**
Bubble	11.4%	20.697
PRoPHETv2	14.5%	25.683
DEBTT [1]	12.1%	**19.146**
DRAFT [2]	14.7%	**25.348**

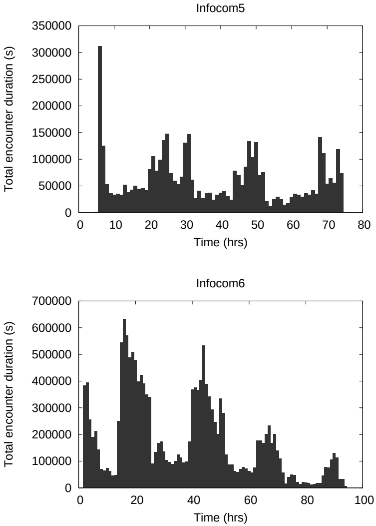
1. Orlinski, M. and Filer, N. (2012) 'Distributed expectation-based spatio-temporal cluster detection for pocket switched networks', IFIP Wireless Days.

2. Orlinski, M. and Filer, N. (2013) 'The rise and fall of spatio-temporal clusters in mobile ad-hoc networks', Journal of Ad Hoc Networks.

Why is data delivery so low?



(a) New encounters per hour



(b) Hourly total encounter durations

Why are overheads so high?

- All of the protocols use message duplication to increase data delivery chance.
- Some messages are duplicated up to 50 times because we can't predict when the next "useful" encounter will take place.

Future work

- Working toward human encounter **forecasting** to predict when encounters will occur.
- We need bigger human encounter datasets that better reflect daily life.

Thank you for listening

Matthew Orlinski
me@shigs.co.uk