Digital Epidemiology: Challenges in Data Collection in Developing Countries

Eiko Yoneki
eiko.yoneki@cl.cam.ac.uk
http://www.cl.cam.ac.uk/~ey204

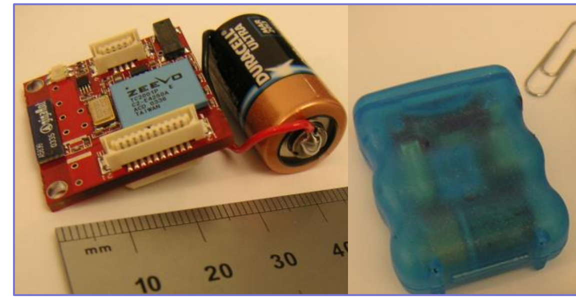
Systems Research Group
University of Cambridge Computer Laboratory

Opportunistic Networks

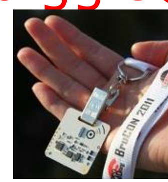
- Pocket Switched Networks: Devices carried by people, thus ‘do what users do’ EU FP7 Haggle Project (2007-2010)

- Pocket switched networks involve consumer devices and incorporate social aspects and opportunistic communication

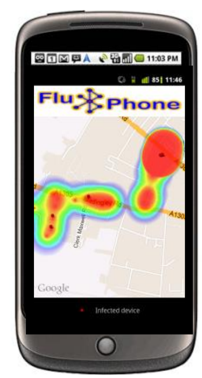
Measure Human Contact Networks



- Sensors
  - Bluetooth Intel iMote
  - 802.15.4 + (magnet, gyroscope)
- RFID Tags 
  - UHF Tag Alien ALN-9640 - "Squiggle®" Inlay
  - OpenBeacon active RFID Tag



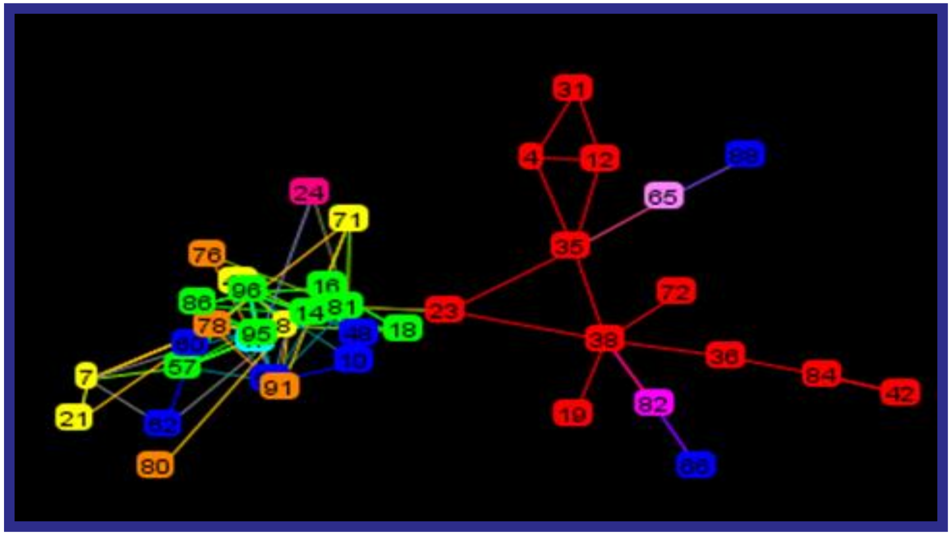
- Mobile Phones
  - Virtual Disease Application Nokia, Android
  - FluPhone Application Nokia, Android
  - AroundYou Application Nokia
  - GPS, Google latitude



- GPS Logger
- Online Social Networks
  - Twitter, Facebook, Foursquare…

Understanding Contact Networks

- Contact networks: Real-world network of time dependent contacts
  - A -> B -> C -> A
- How do we uncover modes of spread?



Spread of Infectious Diseases

- Thread to public health: e.g., 

SARS, AIDS, Ebola
- Current understanding of disease spread dynamics
  - Epidemiology: small scale empirical work
- Real-world networks are far more complex
  - Advantage of real world data
  - Emergence of wireless technology for proximity data
  - Post-facto analysis and modelling yield insight into human interactions

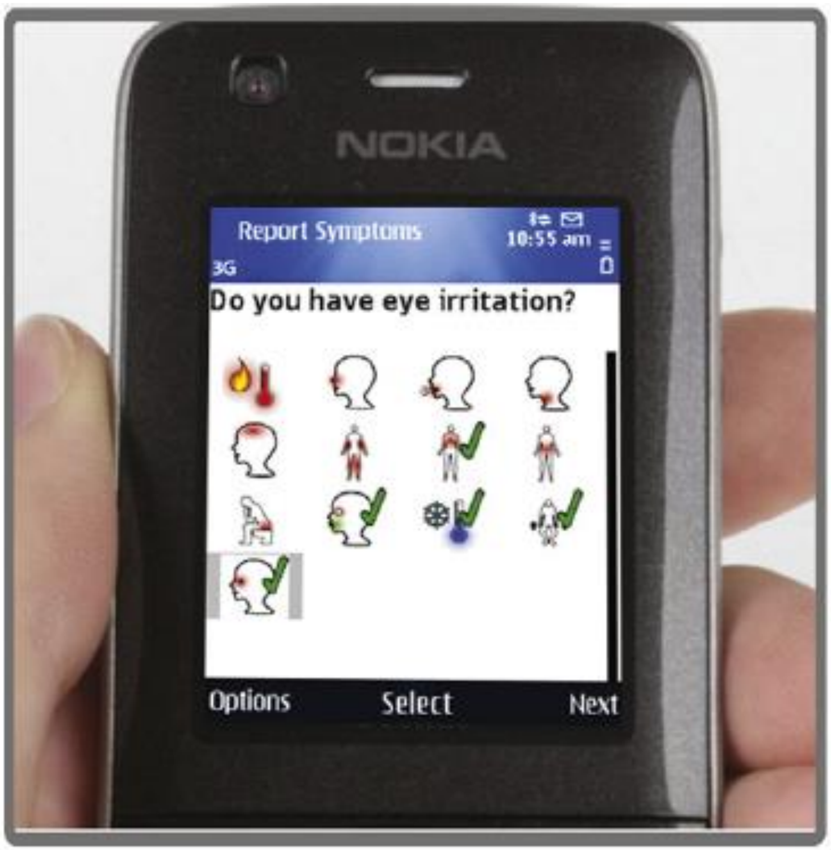
Analogue to Epidemiology

Robust Epidemic Routing

Modelling realistic infectious disease spread/prediction

FluPhone Project

- Scan Bluetooth devices every 2 minutes
- Ask symptoms



FluPhone Project

- Understanding behavioural r[…] disease outbreaks
- Proximity data collection usi[…] general public in Cambridge

https://www.fluphone[…]

FluPhone | UNIVERSITY OF CAMBRIDGE | UNIVERSITY OF LIVERPOOL

Main page | Information | Help | Contact us

FluPhone Study

This is the home page for the FluPhone study. A study to measure social encounters made be[…] their mobile phones, to better understand how infectious diseases, like 'flu, can spread betwee[…]

This study will record how often different people (who may not know each other) come close to[…] part of their everyday lives. To do this, we will ask volunteers to install a small piece of software[…] on their mobile phones and to carry their phones with them during their normal day-to-day activ[…] will look for other nearby phones periodically using Bluetooth, record this information and send[…] research team via the cellular phone data service. This information will give us a much better u[…] often people congregate into small groups or crowds, such as when commuting or through wor[…] activities. Also, by knowing which phones come close to one another, we will be able to work o[…] people actually are, and how fast diseases could spread within communities. We are also aski[…] inform us of any influenza-like symptoms they may experience during the study period, so that[…] spread of 'flu to the underlying social network of encounters made.

BBC Mobile | News Sport Weather iPlayer TV

NEWS CAMBRIDGESHIRE

Home | World | UK | England | N. Ireland | Scotland | Wales | Business | Politics | Health | Education | Sci/En[…]

4 May 2011 Last updated at 17:49

FluPhone app 'helps track spread of infectious diseases'

A mobile phone application could help monitor the way infectious diseases such as flu are spread.

The FluPhone app tracks volunteer 'infected subjects' using Bluetooth technology

The FluPhone app was developed by researchers at the University of Cambridge Computer Laboratory.

Volunteers' phones fitted with the app "talk" to each other, recording how many people each 'infected subject' meets during an imaginary epidemic.

The university is one of seven institutions working on the study to reduce the impact of epidemics.

The FluPhone app uses Bluetooth technology to anonymously record interaction between volunteers involved in the study.

When mobile phones come into close proximity, that fact is recorded and data is sent automatically to the research team.

Related Stories

Web surveillance maps global disease trends

'Valuable insight'

Professor Jon Crowcroft and Dr Eiko Yoneki, co-principal investigators of the study, said they believed the collected data could be used to simulate social interaction during a real epidemic or pandemic.

A three-month FluPhone pilot study, using a basic version of the app, was conducted in Cambridge in 2010.

Dr Yoneki said: "The data was a valuable insight into how human communities are formed, how much time people spend together, and how frequently they meet.

"Such data show complex network-like structures, which is very useful for understanding the spread of disease."

Prof Crowcroft explained epidemiologists traditionally monitor how a disease spreads by asking patients to keep diaries of their movements and social contacts.

Monitoring behaviour during a simulated epidemic could help prevent the disease spreading

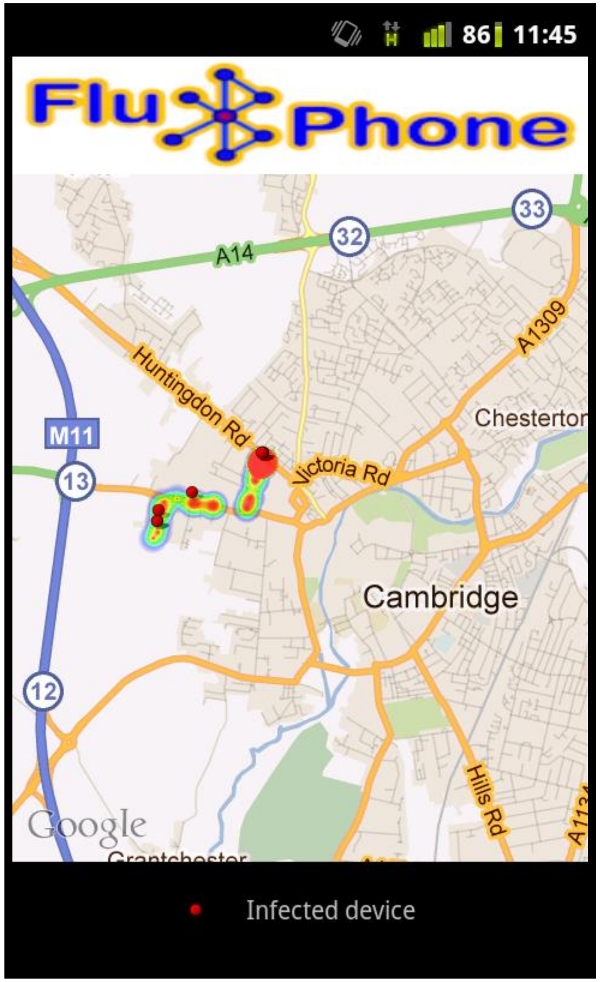
"That's very heavy-going and people often forget to do it, or forget who they've met," he said.

The FluPhone app was, he explained, a more reliable way to record contact between 'infectious subjects'.

"Provided we have people's permission, we can upload the data, and medical researchers can see who met whom within the set of volunteers, without there being any missing encounters.

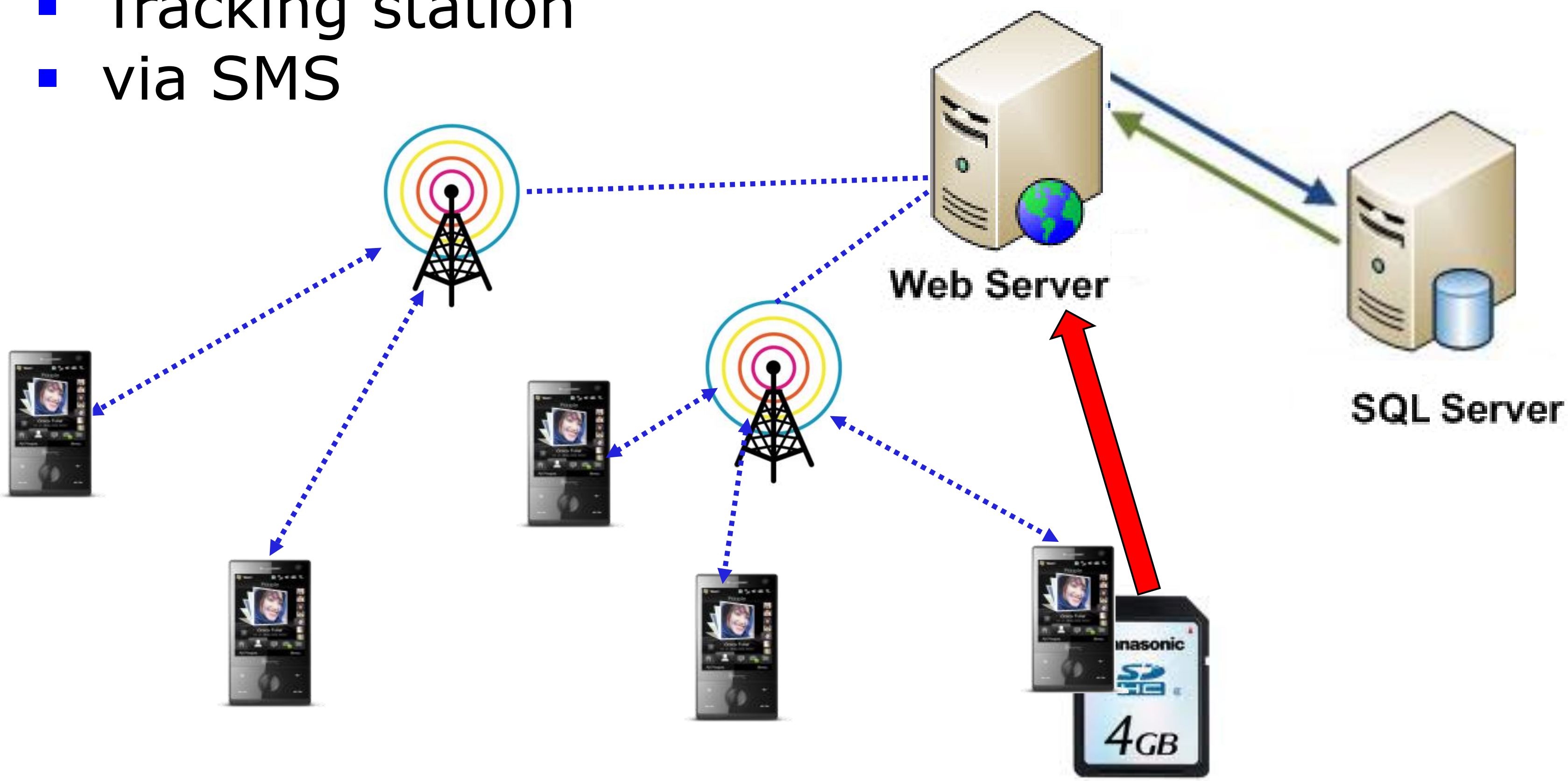
Trajectory of Encountering Sick People

- Integration with GPS equipped Smartphones



FluPhone – Data Collection

- via GPRS/3G FluPhone server collects data
- Uploading via Web
- via memory card
- Tracking station
- via SMS



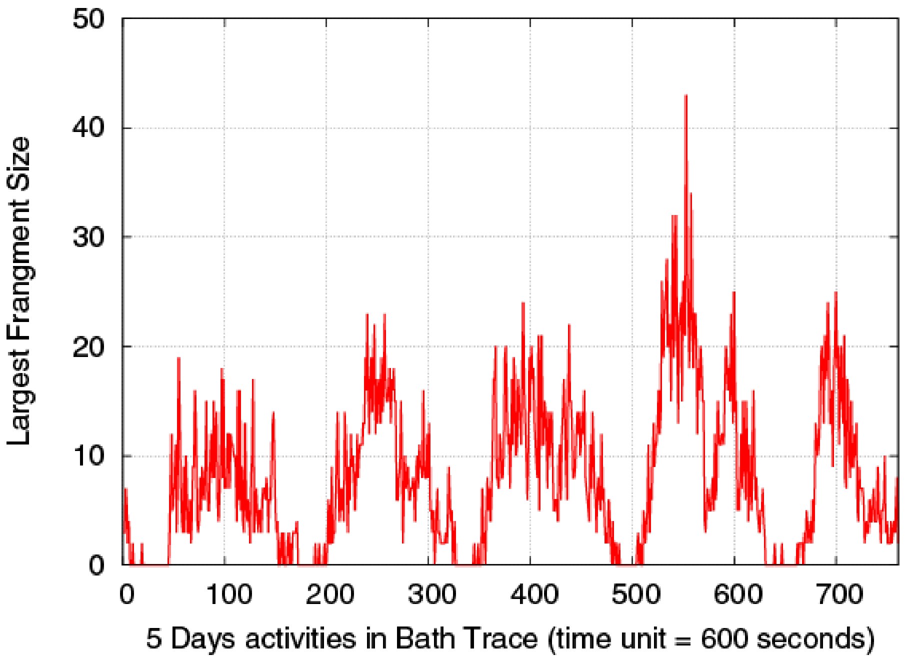
Trace Data of Contact Networks

- Analysis of dynamic network structure
  - How does community structure affect epidemic spread?
  - How do hubs influence temporal or spatial effects, and how does this affect the transmission characteristics of disease?

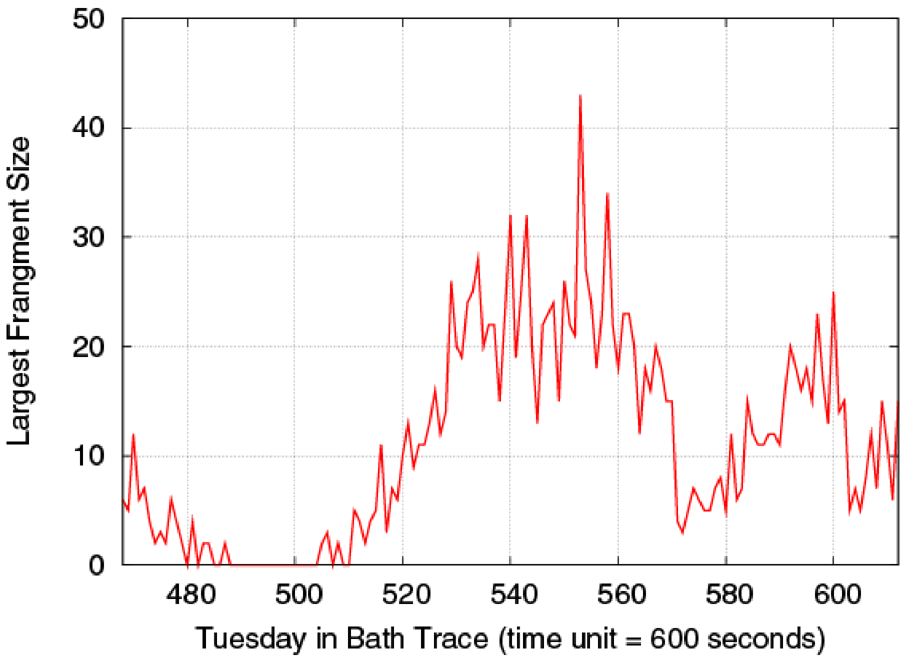
Experimental data set	Cambridge	Infocom06	MIT
Device	iMote	iMote	Phone
Network type	Bluetooth	Bluetooth	Bluetooth
Duration (days)	11	3	246
Granularity (seconds)	600	120	300
Number of Devices	36	78	97
Number of contacts	10,873	191,336	54,667
Average # Contacts/pair/day	0.345	6.7	0.024

Regularity of Network Activity

- Size of largest connected nodes shows network dynamics



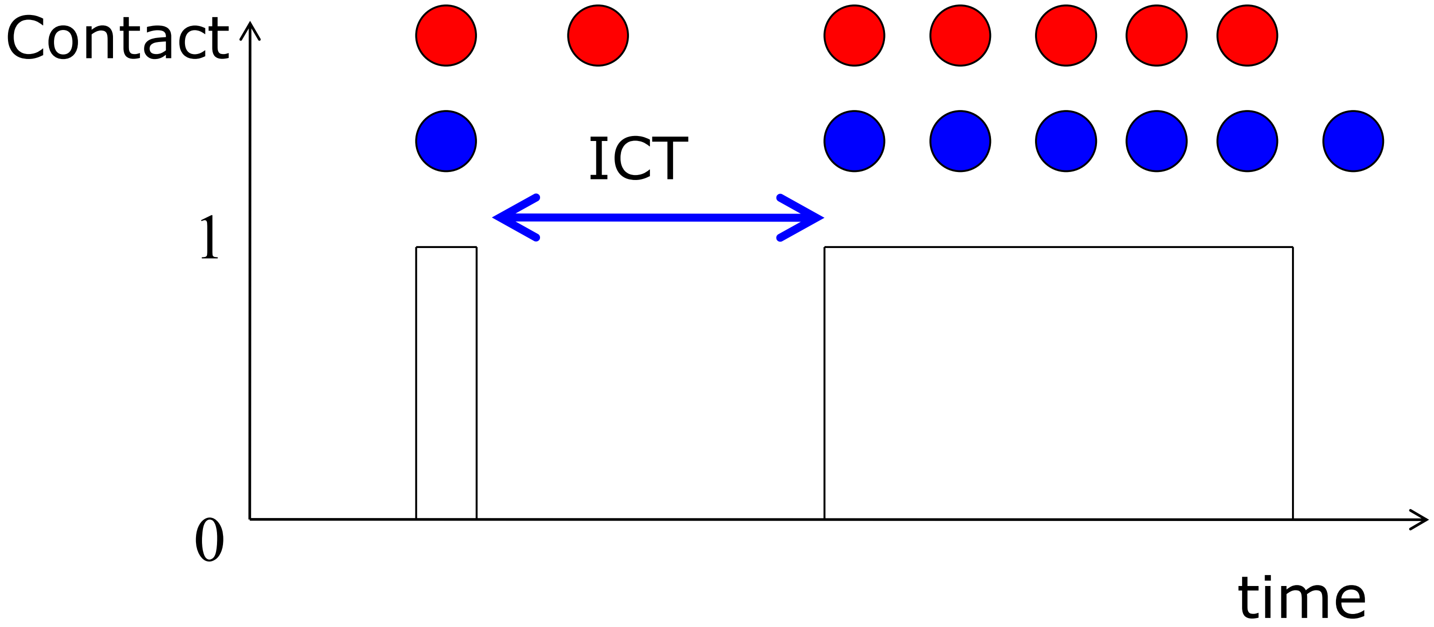
5 Days



Tuesday

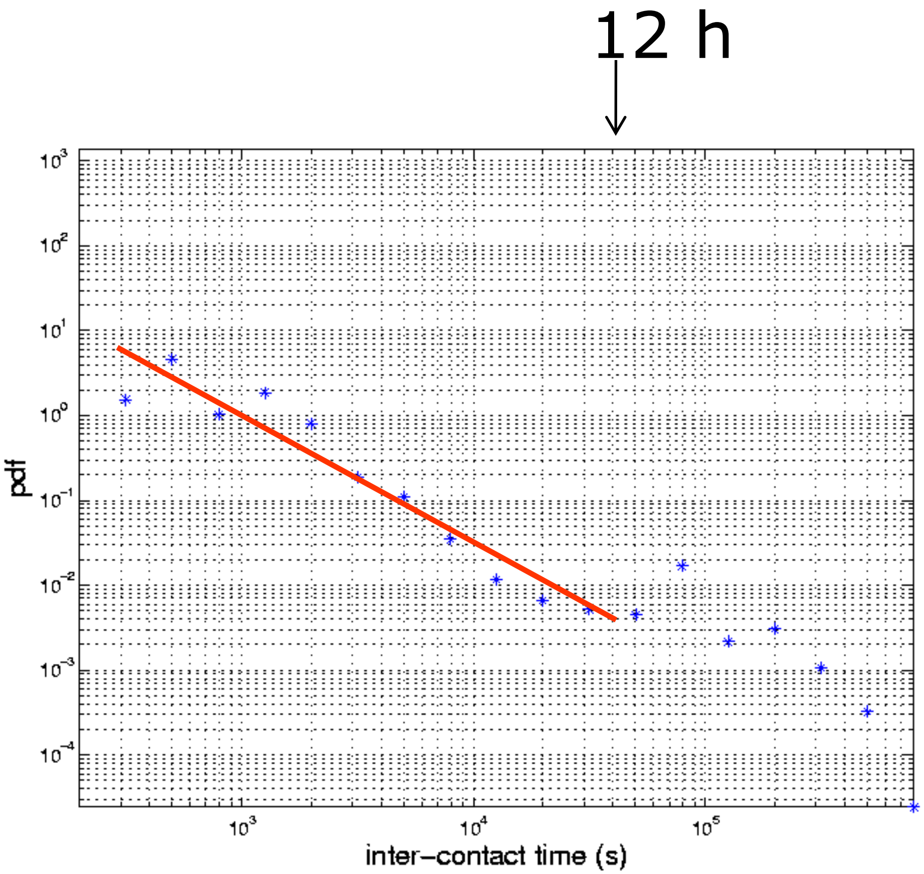
Inter-Contact Time (ICT)

- Calculated all possible inter-contact times between any two nodes, where ICT is defined as the time between the end of contact between two nodes and the start of next contact between the same two nodes

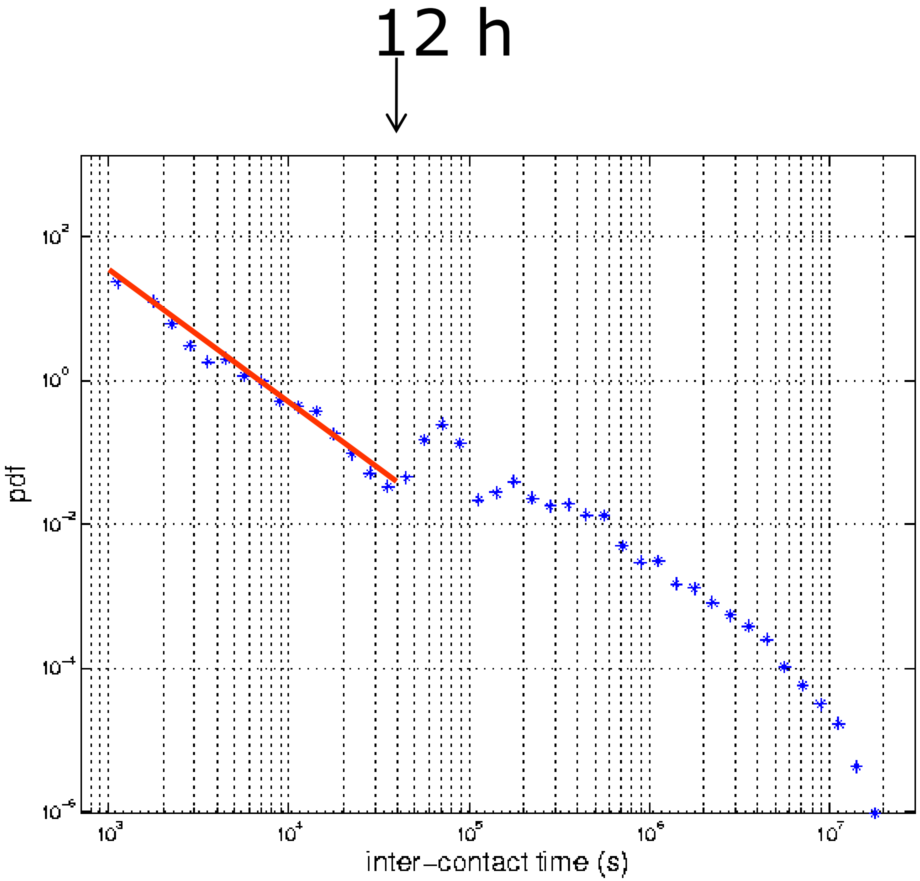


ICT: Random and Scale-free

- Sufficiently short time scales (<12 hours): ICT dist is approximated by power law



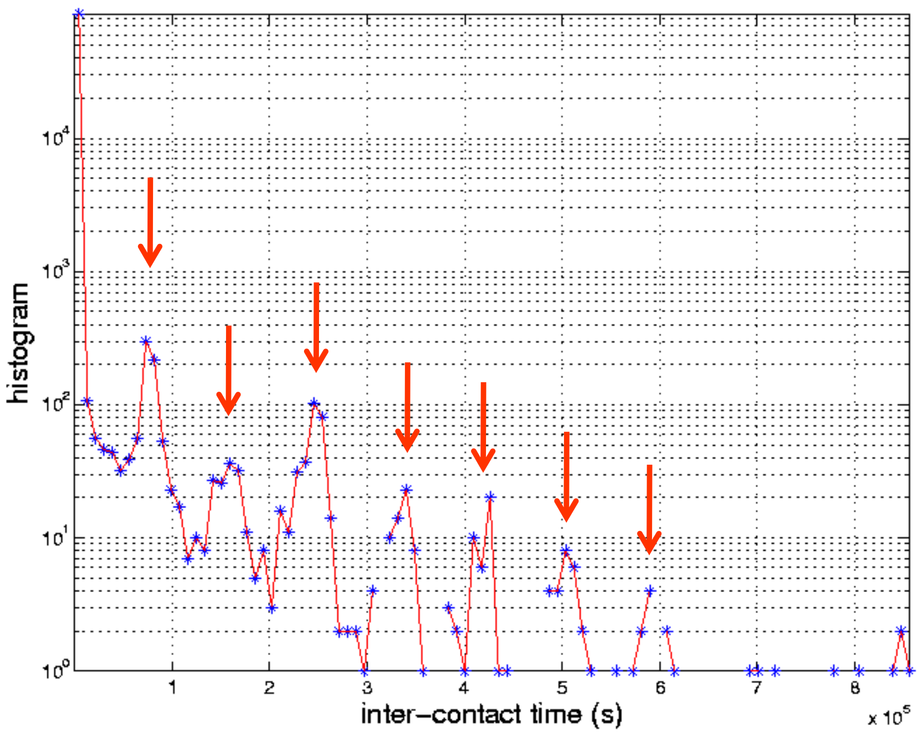
Cambridge



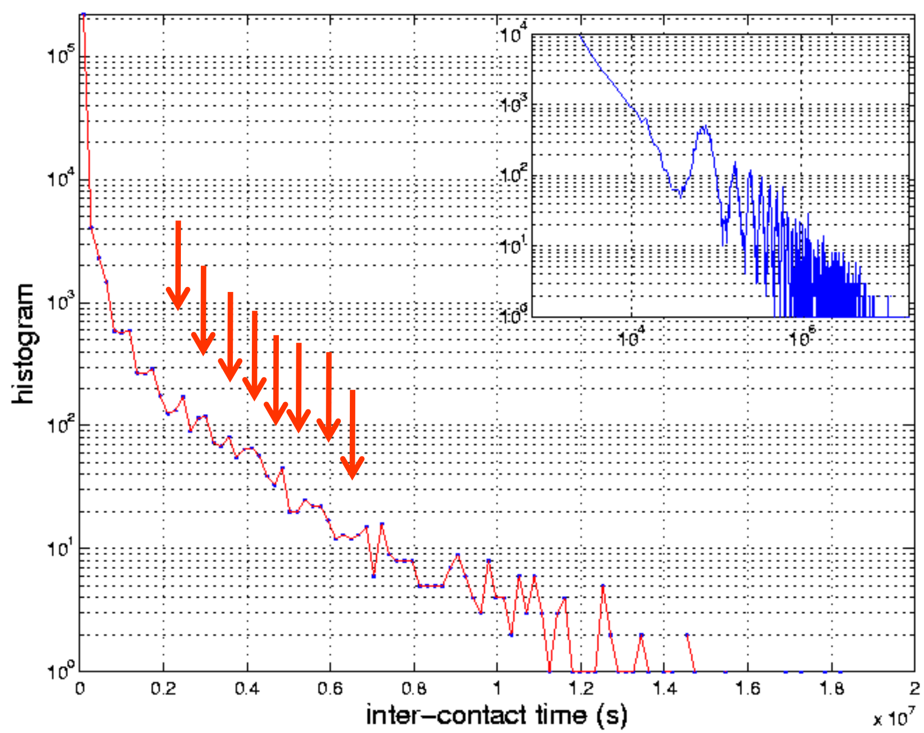
MIT

ICT: Periodic

- Environmental, biological, and social constraints may have rhythms that encourage repeated encounters such as the daily to-ing and fro-ing between work and home. This gives ICT separated by 24 hours



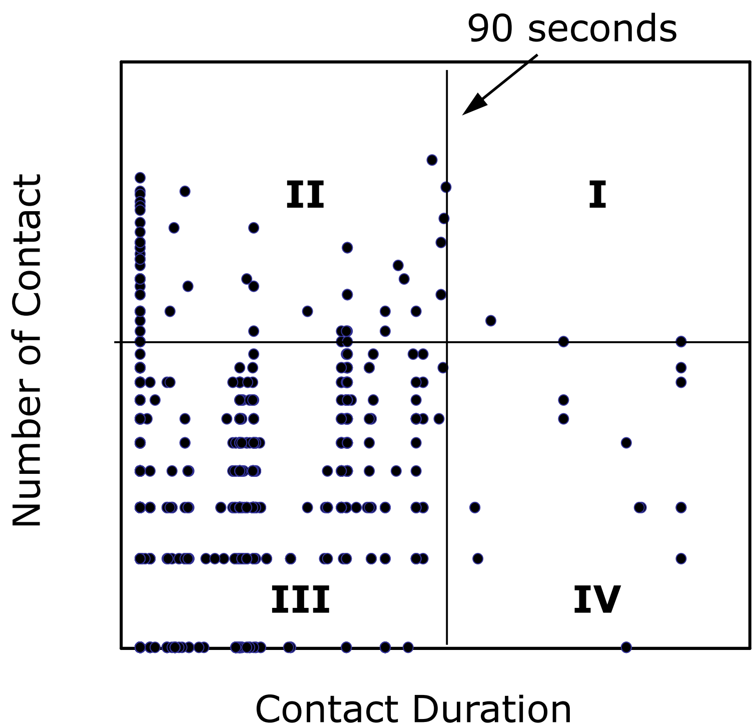
Cambridge



MIT

Edge Weight

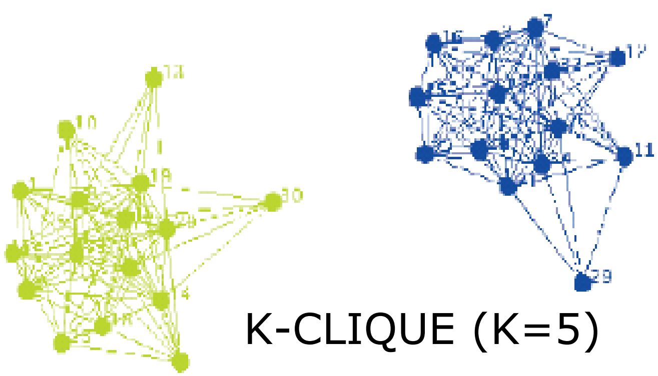
I. High Contact Nº - Long Duration: Community
II. High Contact Nº - Short Duration: Familiar Stranger
III. Low Contact Nº - Short Duration: Stranger
IV. Low Contact Nº - Long Duration: Friend



Uncovering Community

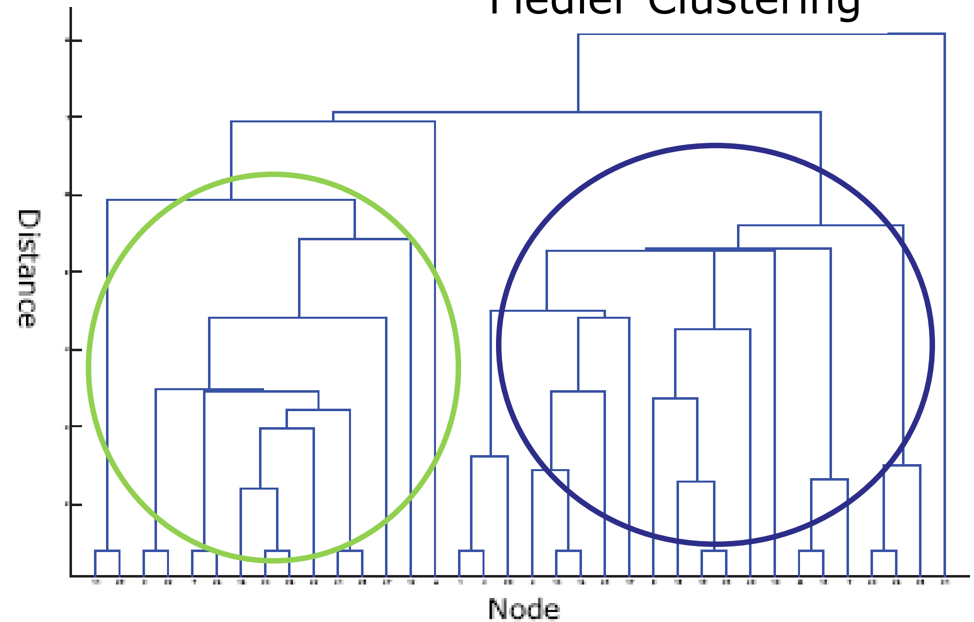
- Contact trace in form of weighted (multi) graphs
  - Contact Frequency and Duration
- Use community detection algorithms from complex network studies
  - K-clique [Palla04], Weighted network analysis [Newman05], Betweenness [Newman04], Modularity [Newman06], Fiedler Clustering etc.

Cambridge Trace



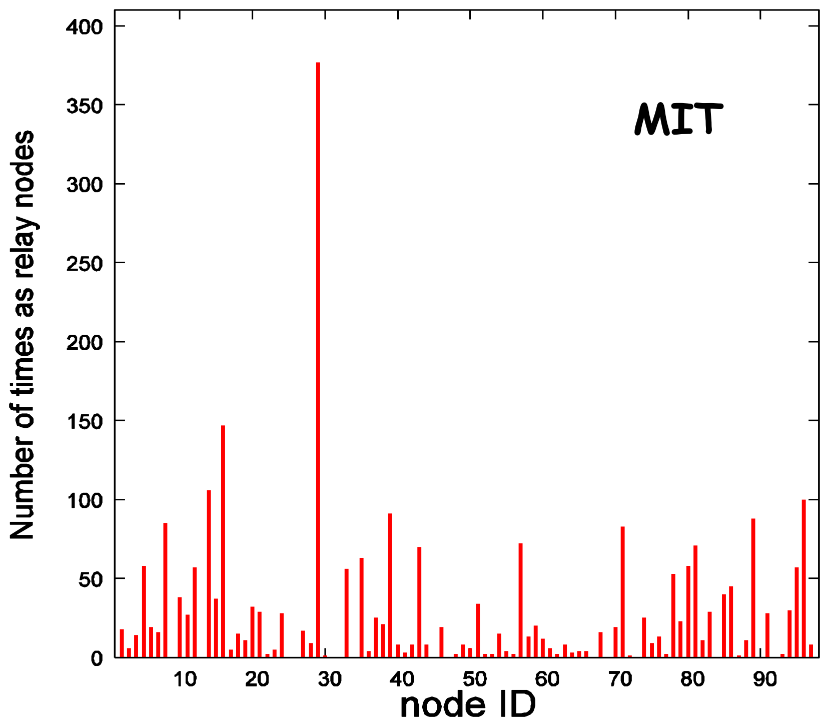
K-CLIQUE (K=5)

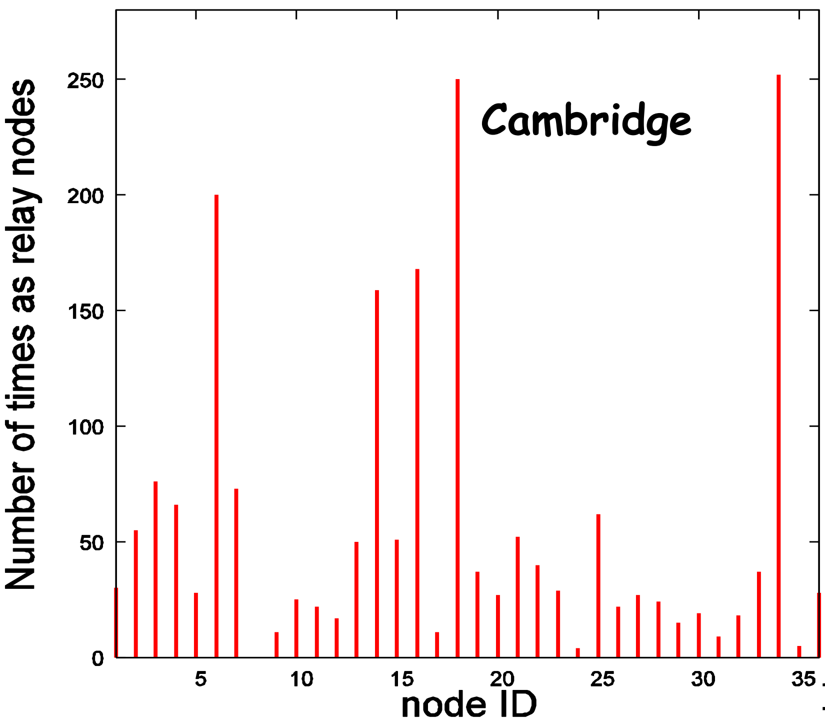
Fiedler Clustering



Betweenness Centrality

- Frequency of a node that falls on the shortest path between two other nodes
- High ranking nodes ~= Popular nodes

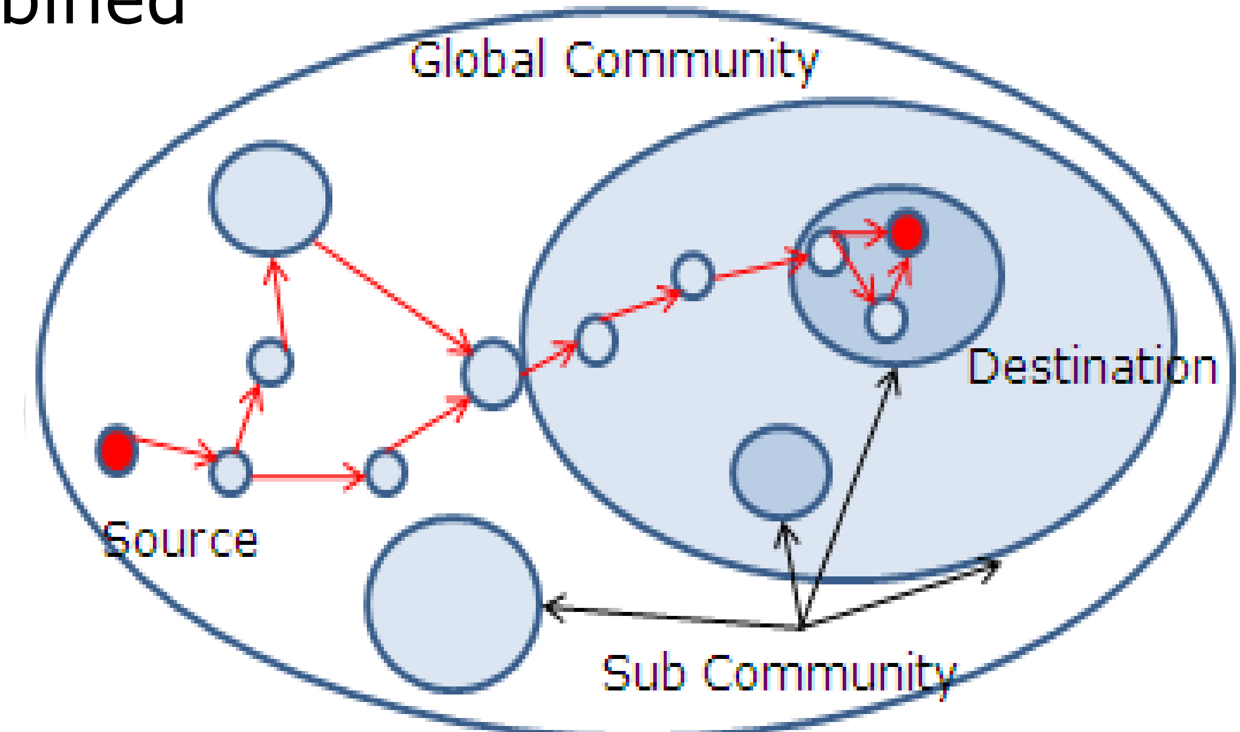




Social Structure for Communication

- Community and Centrality based forwarding

- LABEL Community based
- RANK Centrality based
- BUBBLE Combined

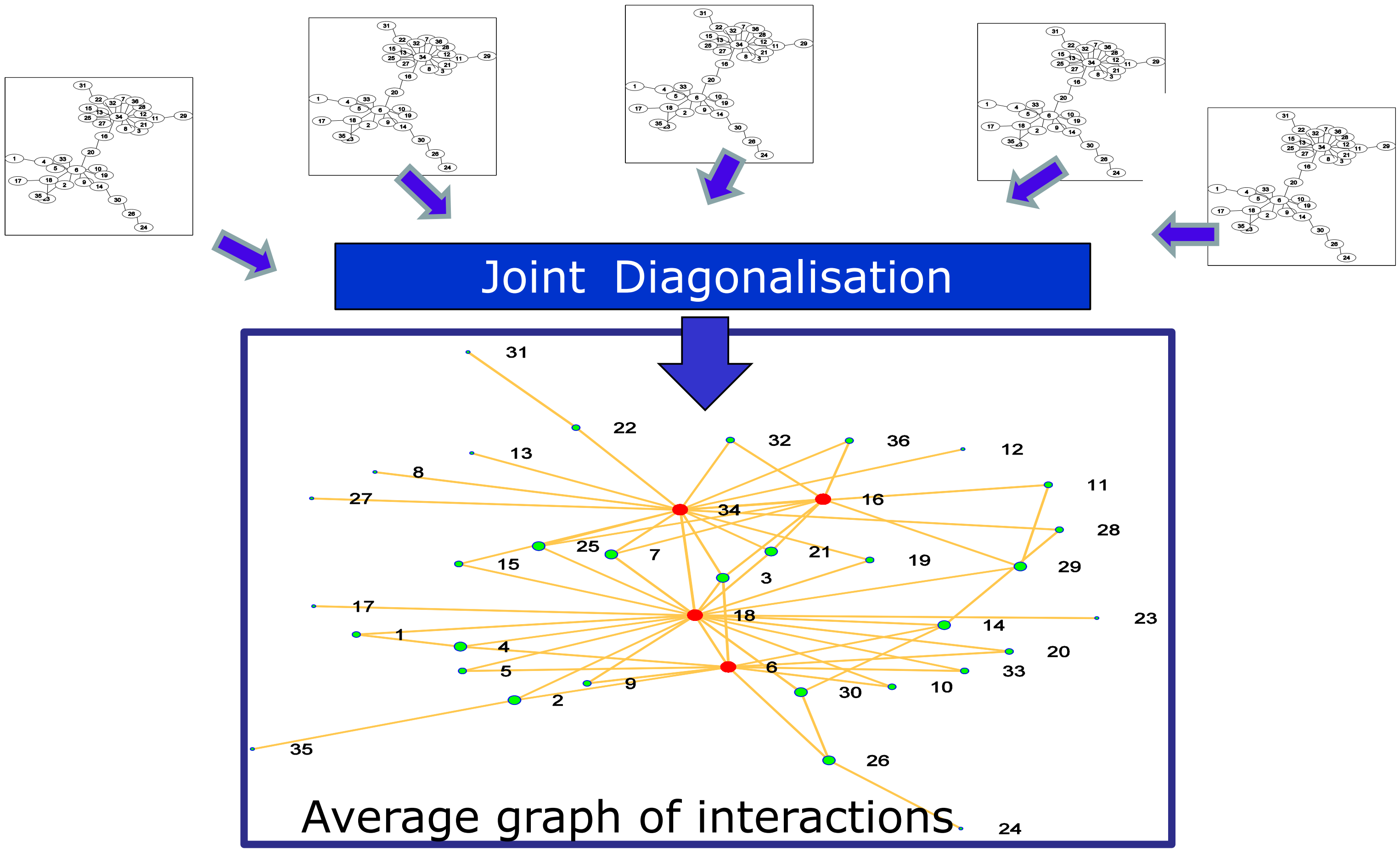


Multiple Spread Modes

- Typical approach: Cluster nodes to build single network or multiple networks within the sliced time windows
  - Aggregate the number of contacts
  - Ignores time
  - Ignores correlation between links
- Solution: Use spanning tree based samples of a network
  - Akin to spreading a disease in the population and recording the order of infection
  - Define an eigen-space average across these trees
  - Distribution of deviations gives the required groups

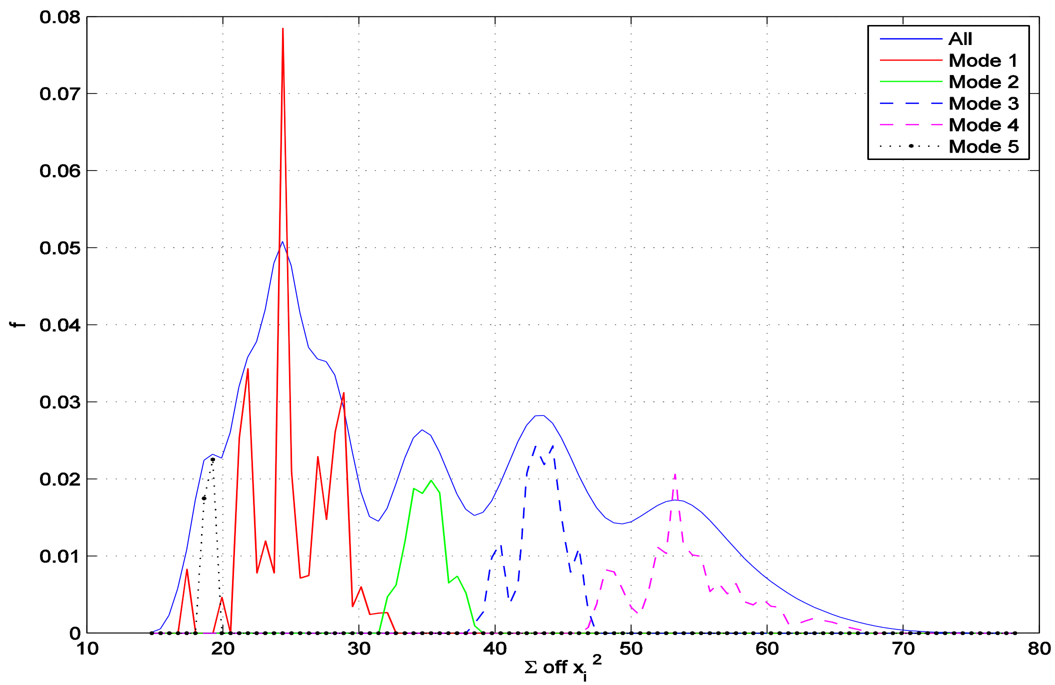
Joint Diagonalisation

- Build by combining many of spanning tree based samples of a network using Joint Diagonlisation → Average Interaction Network



Multiple Network Modes

- Define deviation from the average eigen-space as the sum of off-diagonal elements
- Use Gaussian mixture model for mode determination
- Distribution of deviation from average graph is multi-modal ➡ different behaviour of network

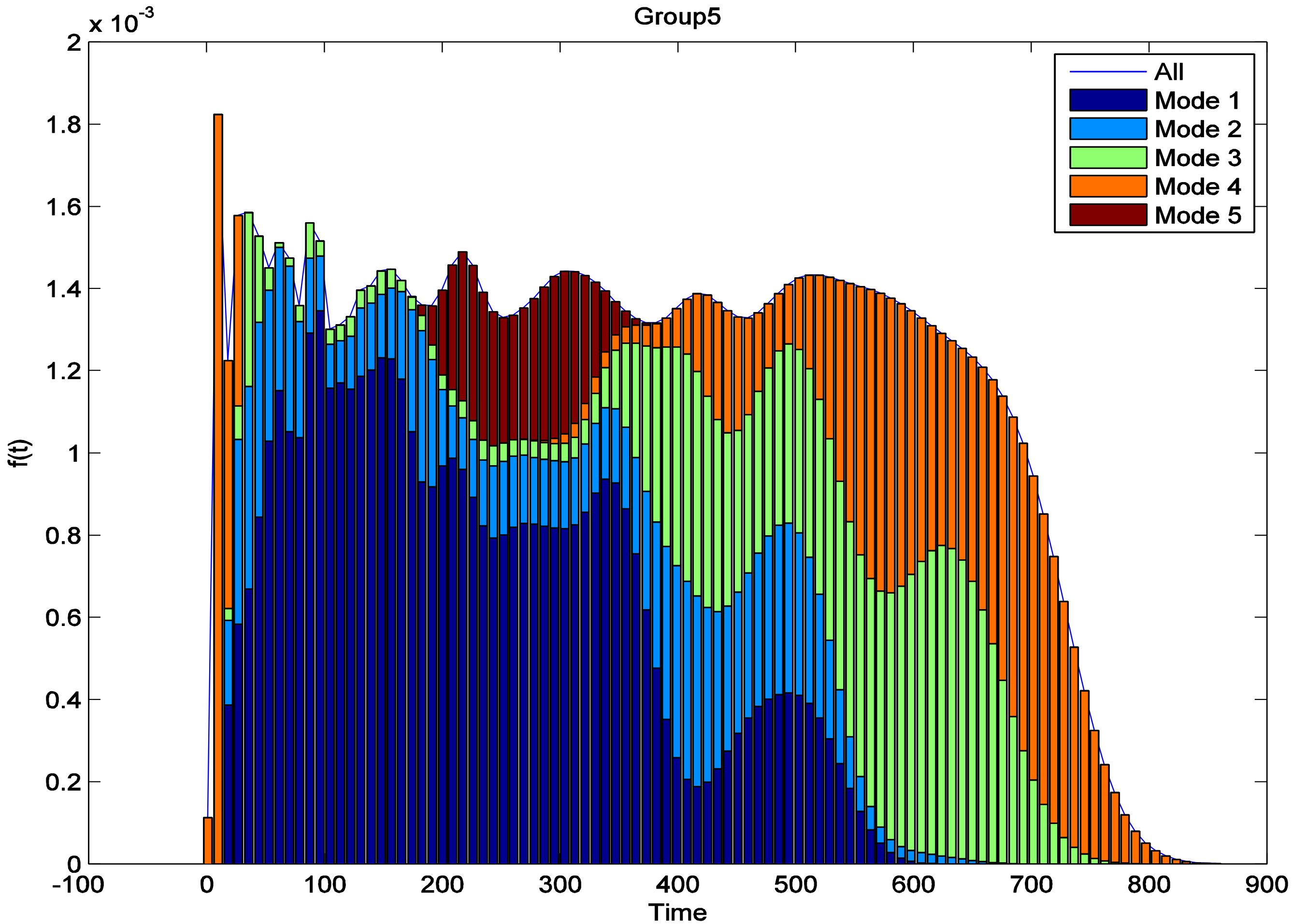


Distribution of deviation (Cambridge data)

Note: A random network shows only one mode

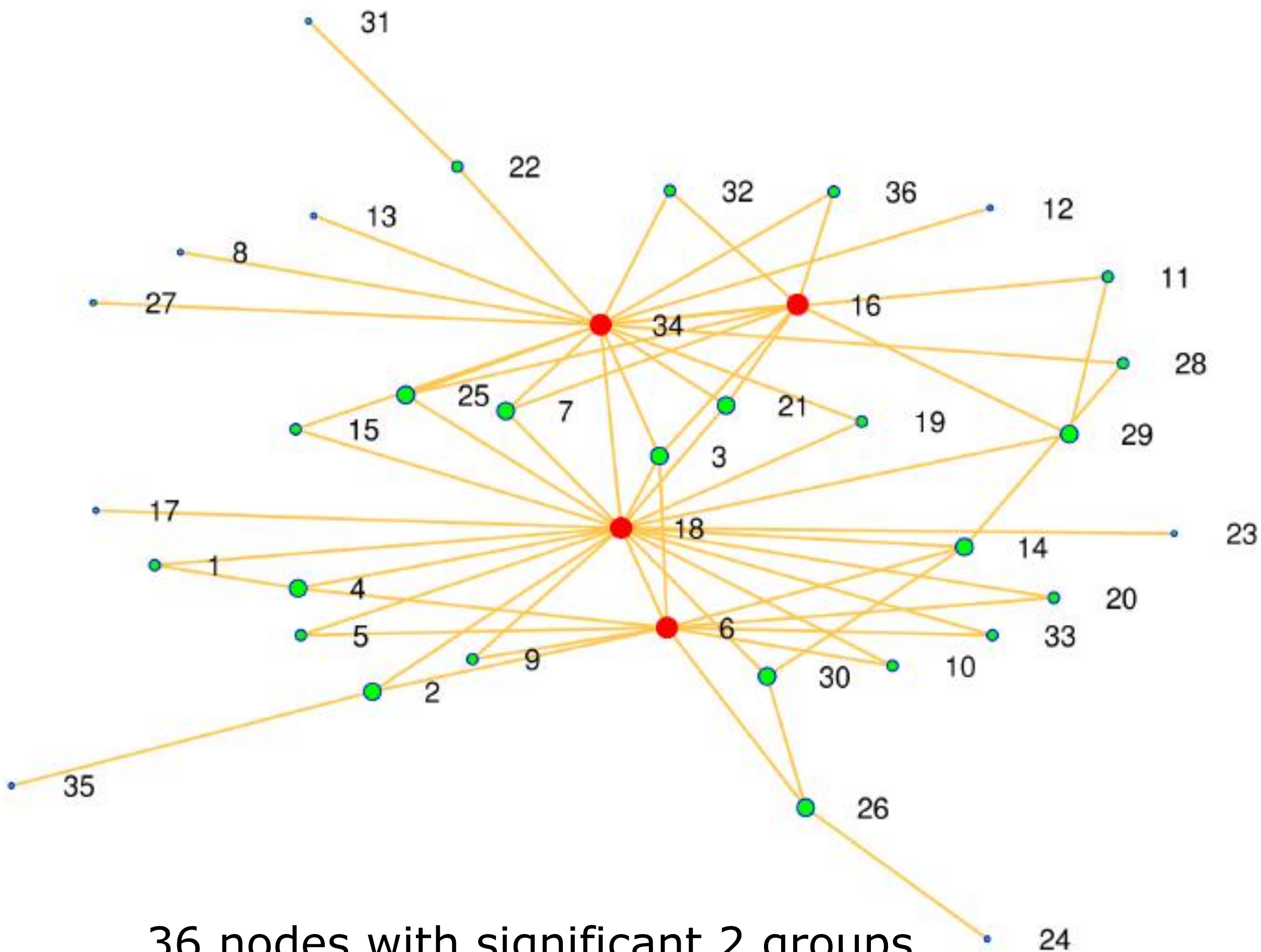
Extract Spread Modes

- Change of mode corresponds with state transition



Distribution of times by mode

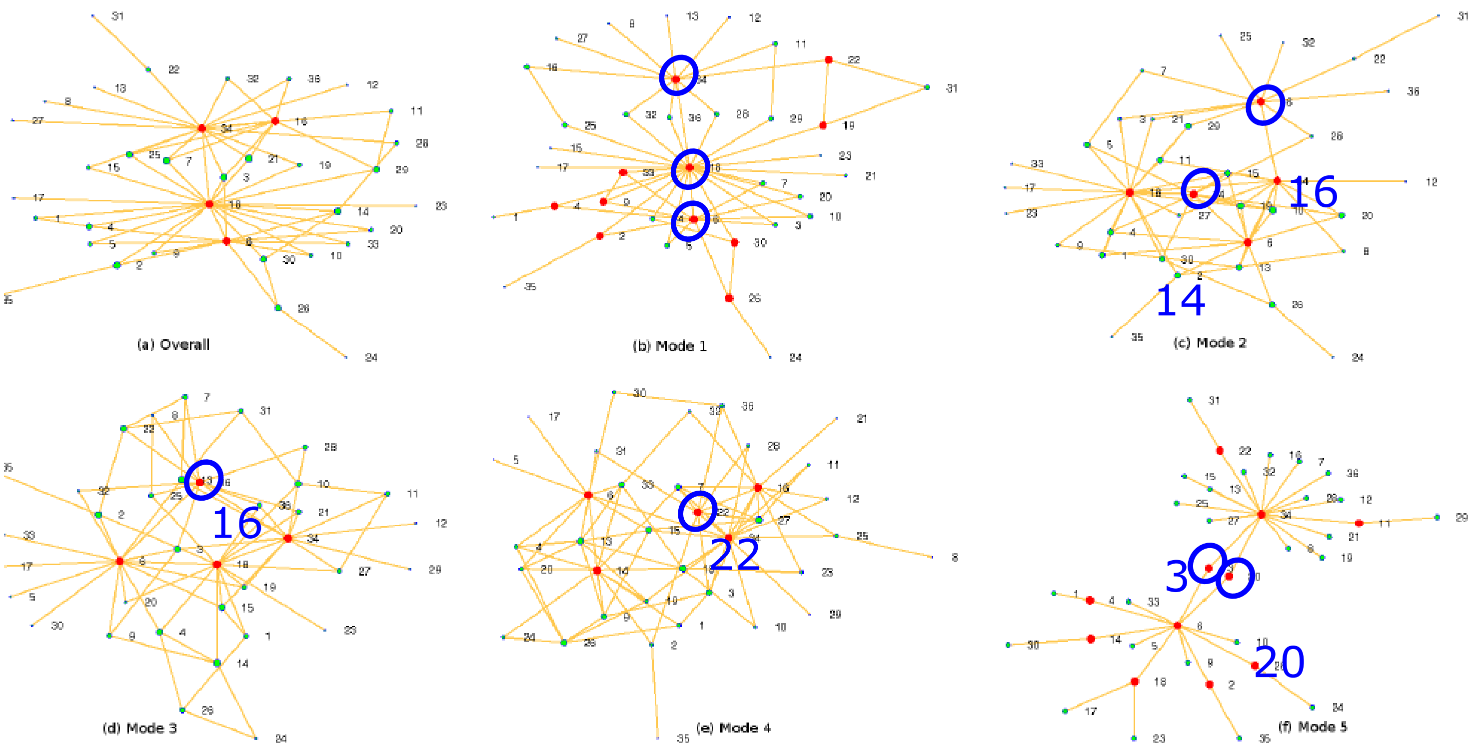
Average Graph of Interactions



36 nodes with significant 2 groups

Network Structure of Each Mode

- Mode 1 shows a highly structured network corresponding to the day when the groups are well defined by group dependent activity
- Mode 5 is particularly interesting as there is an obvious bridge formed by nodes 3 and 20

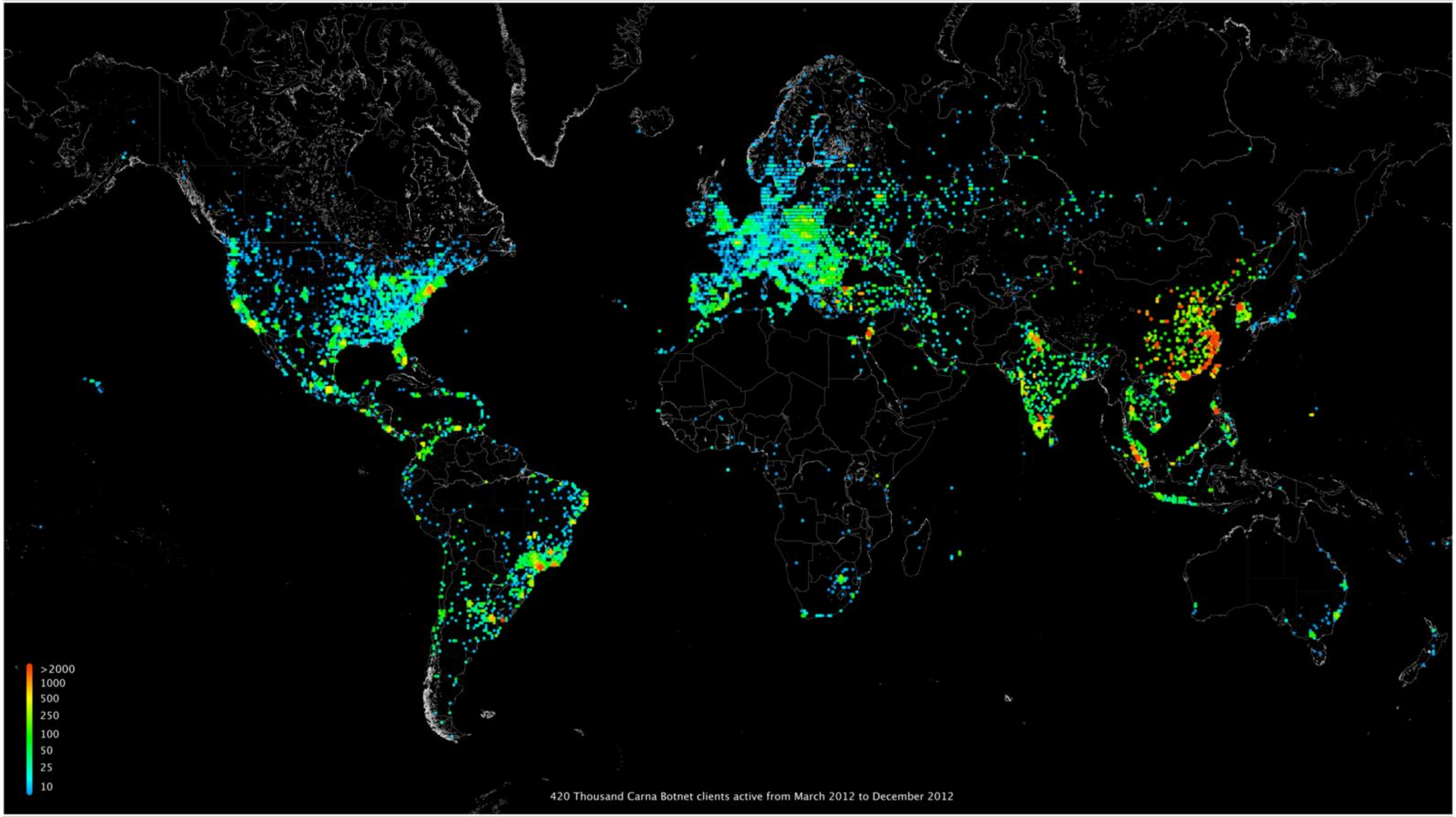


(a) Overall (b) Mode 1 (c) Mode 2 (d) Mode 3 (e) Mode 4 (f) Mode 5

Data Collections in Developing Counries

- Possible study in Africa and South America
- Measles, tuberculosis, meningococcal, respiratory syncytial virus and influenza
- Support various proximity sensing techniques
  - Collect medical symptoms
  - Capture surrounding context (e.g. temperature, light, humidity, GPS-location)
- Combine diary and interview Survey
- Need to be repeated data collection
- →Input for effective vaccination strategies within limited budget in developing countries

Internet Map in the World



Sensing Platform in Remote Region

- Build a platform for sensing and collecting data in developing countries
  - e.g. OpenBeacon Active RFID tags based contact network data collection
  - Build a standalone network for data collection and communication using Raspberry Pi → RasPiNET
  - Inexpensive network setting
  - Support streaming model

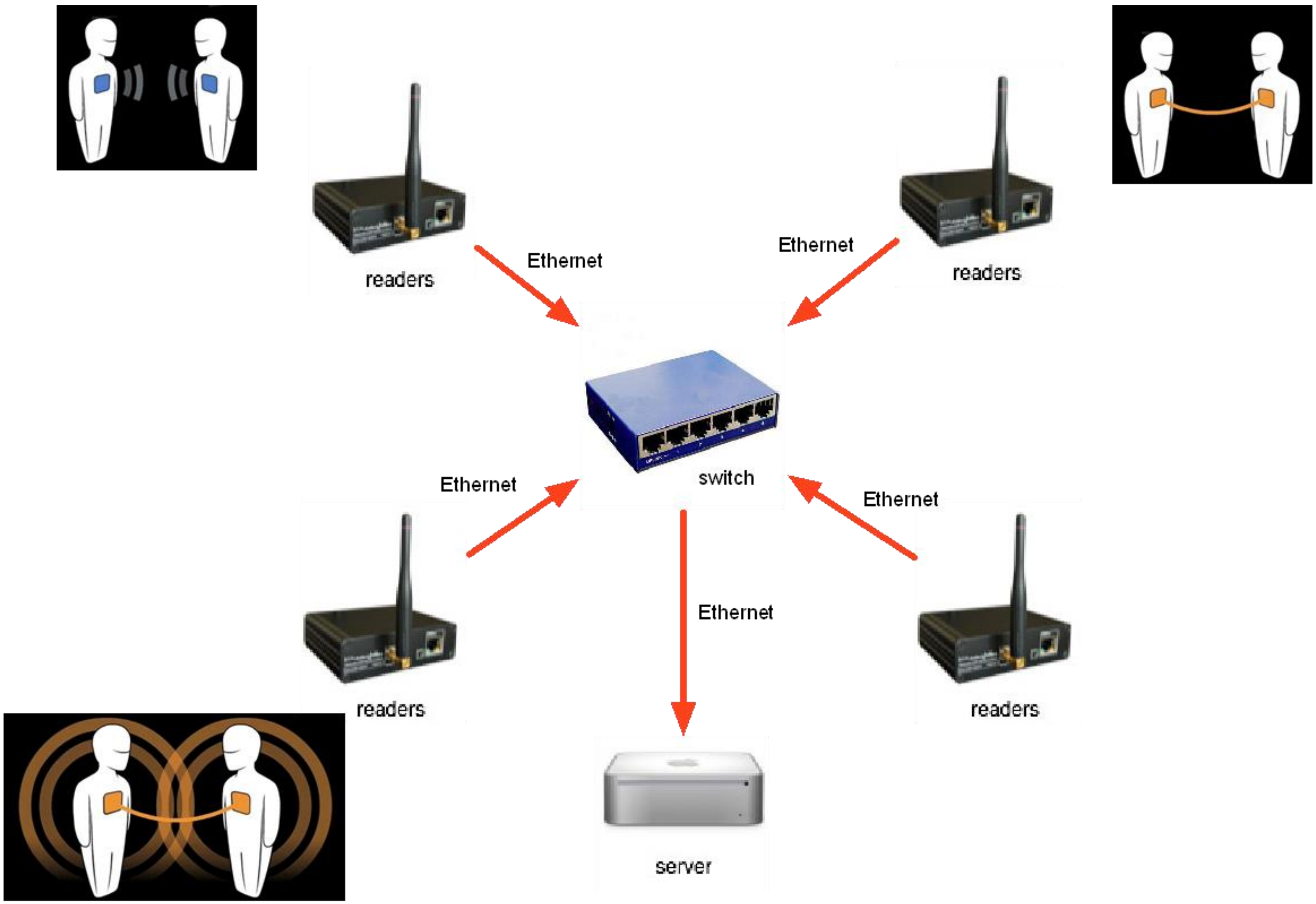
OpenBeacon RFID Tags

- OpenBeacon Active RFID Tags
- Bluetooth has an omnidirectional range of ~10m
- OpenBeacon active RFID tags: Range ~1.5m and only detect other tags are in front of them
- Low Cost ~=10GBP
- Face-to-Face detection
- Temporal resolution 5-20 seconds
- On-board storage (up to ~4 logs)
- Battery life ~2-3 weeks

An OpenBeacon RFID tag

OpenBeacon Ethernet EasyReader

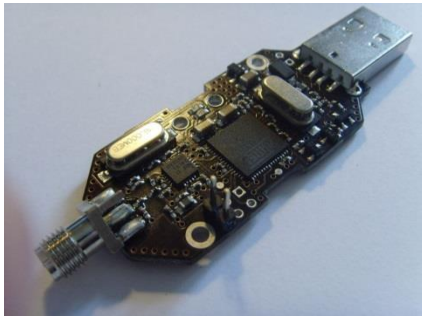
RFID Tag with Ethernet Readers



Raspberry Pi based Reader

- OpenBeacon Ethernet Readers need Ethernet connection (Cannot be deployed outside)

→ Using USB based reader with Raspberry Pi

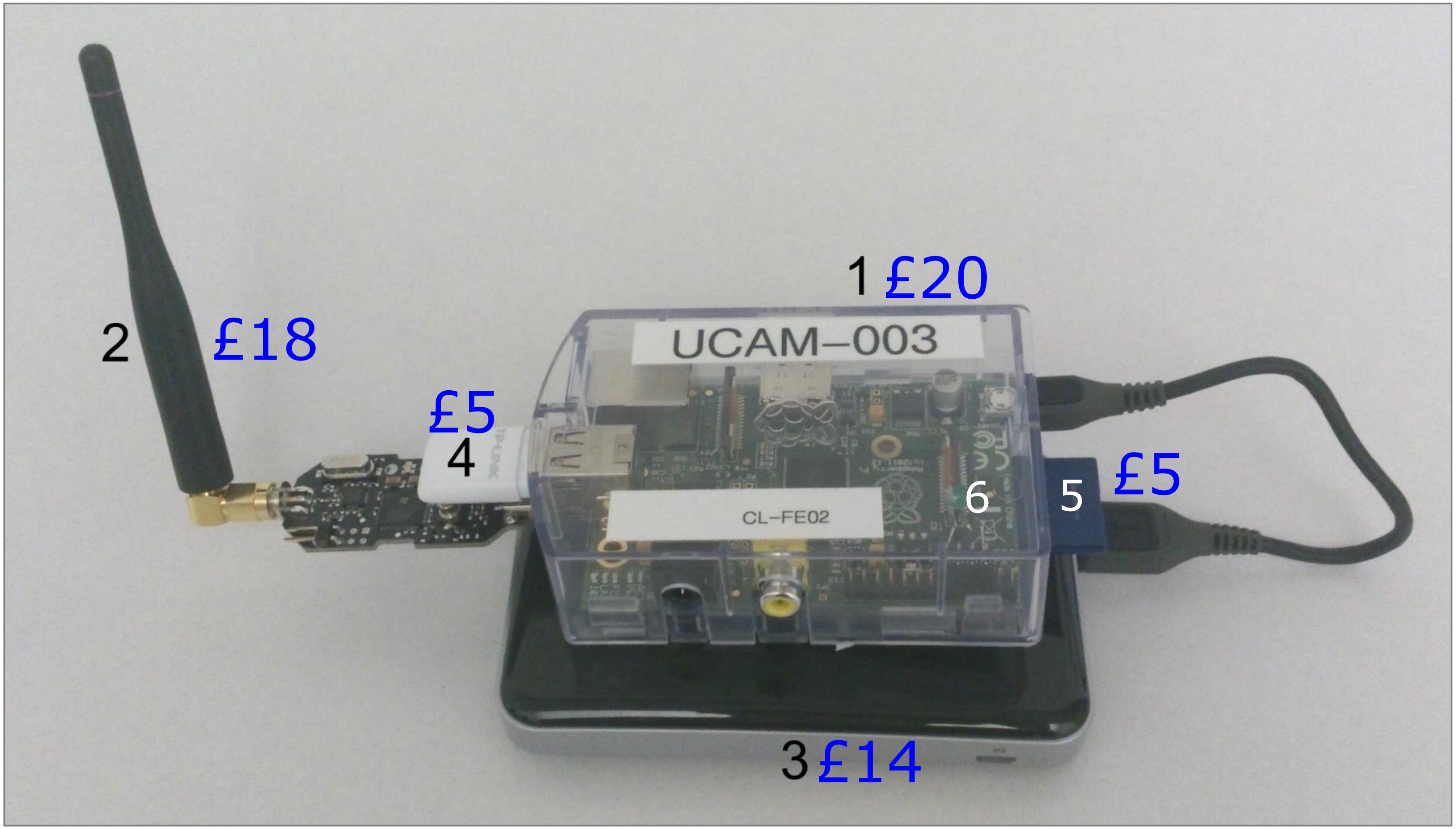


OpenBeacon USB Reader

- USB reader + Raspberry Pi
  - Raspberry Pi (700MHz ARM11 CPU, 512MB RAM, 2 USB ports, SD card port, Ethernet port)
  - WiFi connectivity
  - Mobility (w/ battery pack)
  - Work without a server – SD card storage

TP-Link TL-WN723N
Wifi Dongle

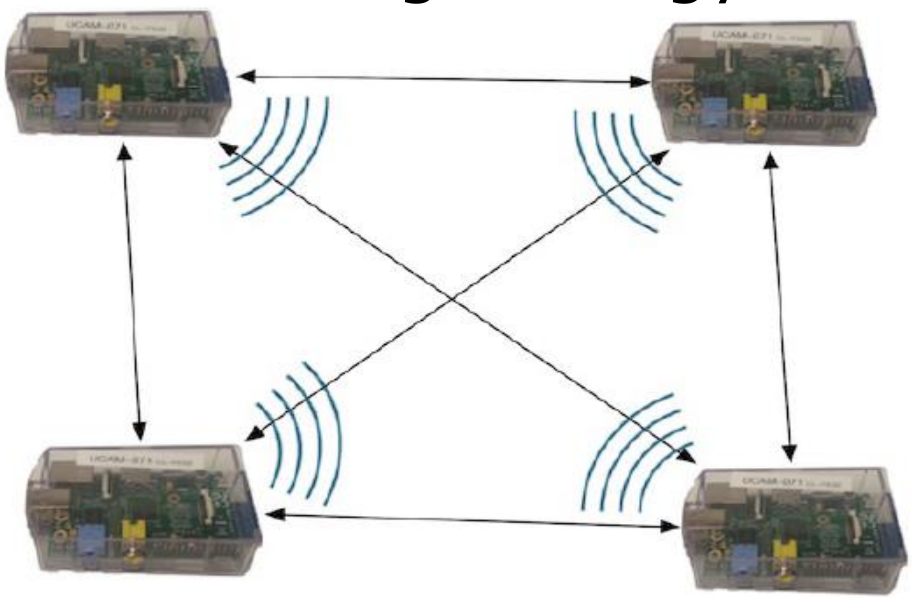
Raspberry Pi OpenBeacon Reader



1. Raspberry Pi
2. OpenBeacon USB reader
3. Battery Pack (7000mAh)
4. WiFi dongle 5. SD Card 6. LED

Option for WiFi Configuration

- Use USB WiFi dongle to setup WiFi Adhoc communication → High energy consumption

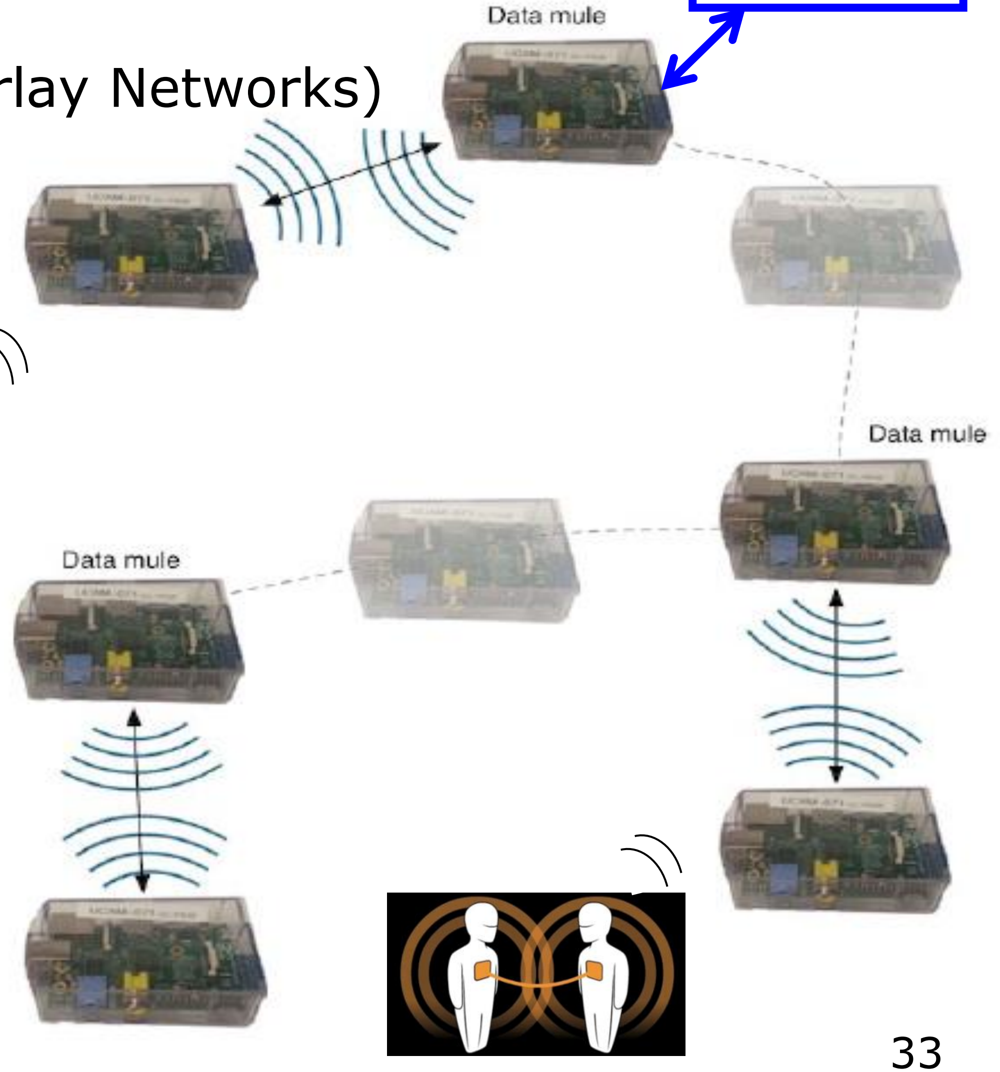


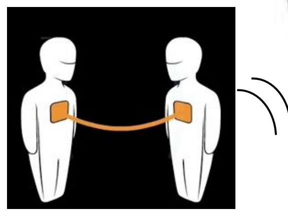
- Software Access Point: One Raspberry Pi acts as Access Point and the others as Clients
- WiFi Direct: All the devices can communicate each other → Reducing energy consumption

DTN in Raspberry Pi

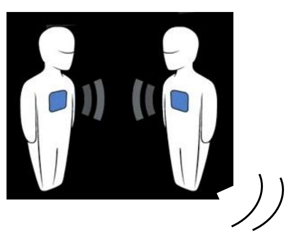
Gateway RasPi with MongoDB

- Bundle Protocol Options
  - DTN2 by DTNRG
  - ION (Interplanetary Overlay Networks)
  - IBR-DTN





- RasPiNET with DTN2
- Software AP based Data Mule



Raspberry Pi based Sensing Platform

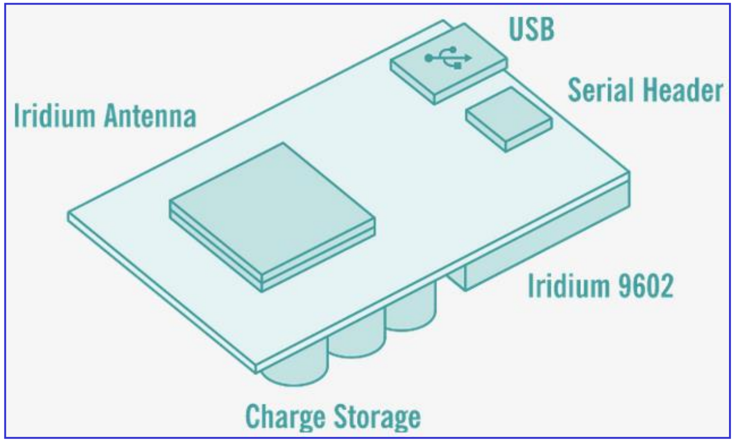
Data can be stored in SD card, transmitted to Data Mule node, or use of WiFi AdHoc mode transmission to Gateway

Configuration: WiFi AdHoc, Software Access Point, WiFi Direct, DTN Data mule

Single/Multiple Satellite Gateway nodes

Satellite Communication

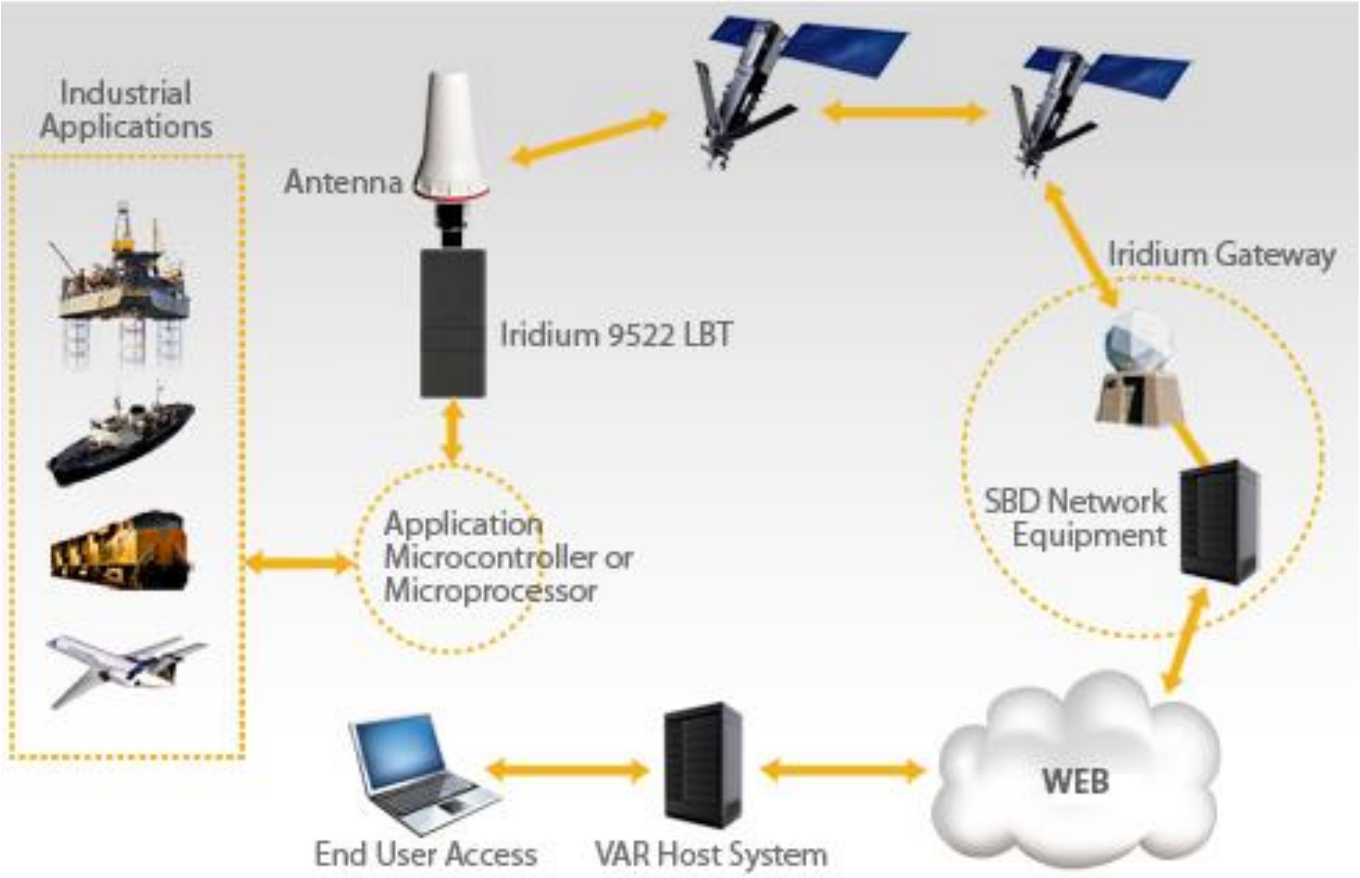
- Satellite module integration in Raspberry Pi
  - RockBLOCK Satellite Module (~=£120)
  - Uses Iridium Satellite Network: Short Burst Data(SBD)
  - Iridium SBD session roughly every 10 seconds
  - To email address, or own web service (i.e. HTTP POST)
  - pay-as-you-go – 34 bytes per message (Hex encoded)
    - 50 credits - 12p/message
    - 20000 Credits – 4p/message



RockBLOCK satellite module

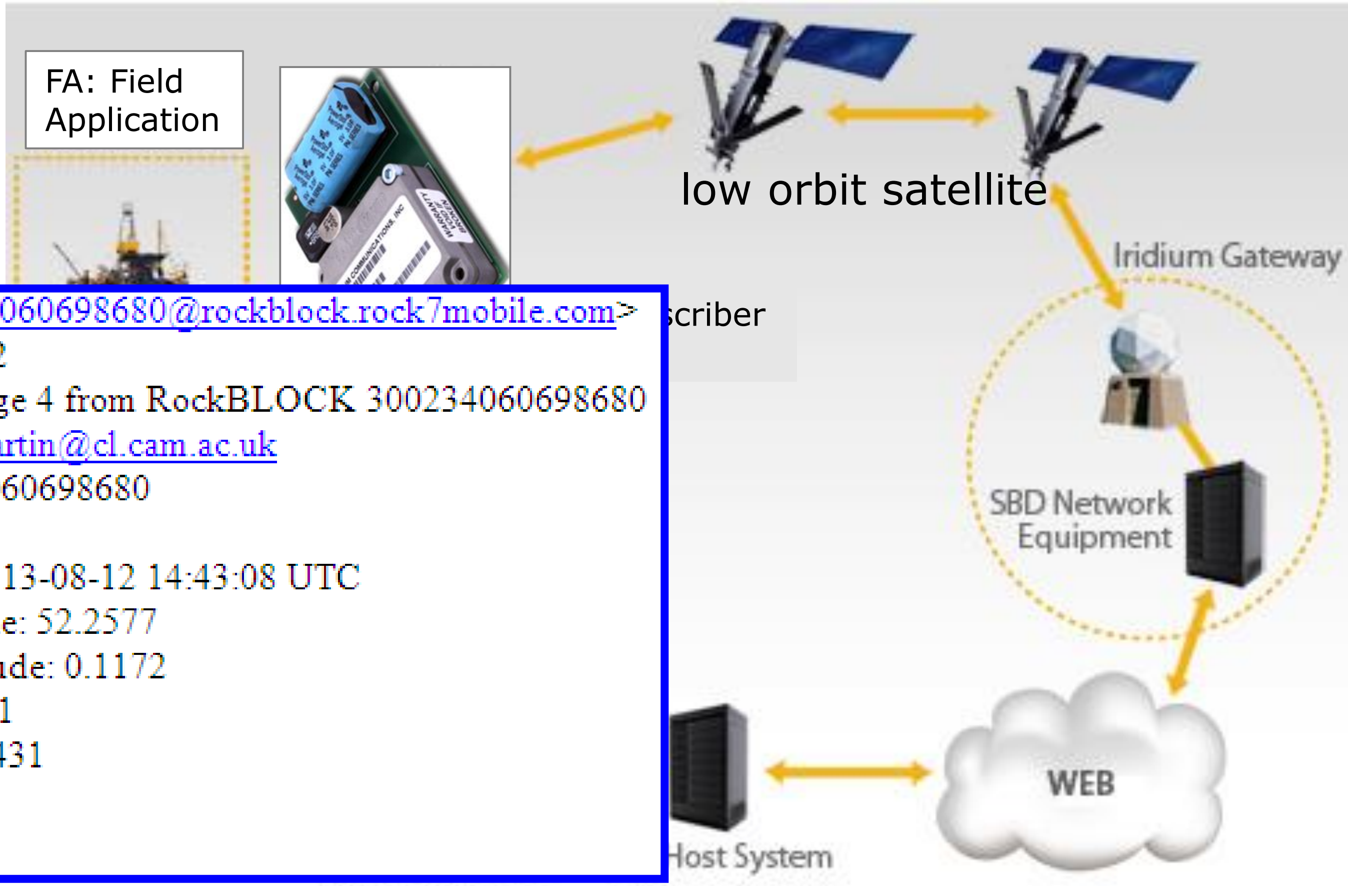
RasPiNET with Satellite Communication

- Satellite module is integrated
- Useful in developing country



Iridium SBD

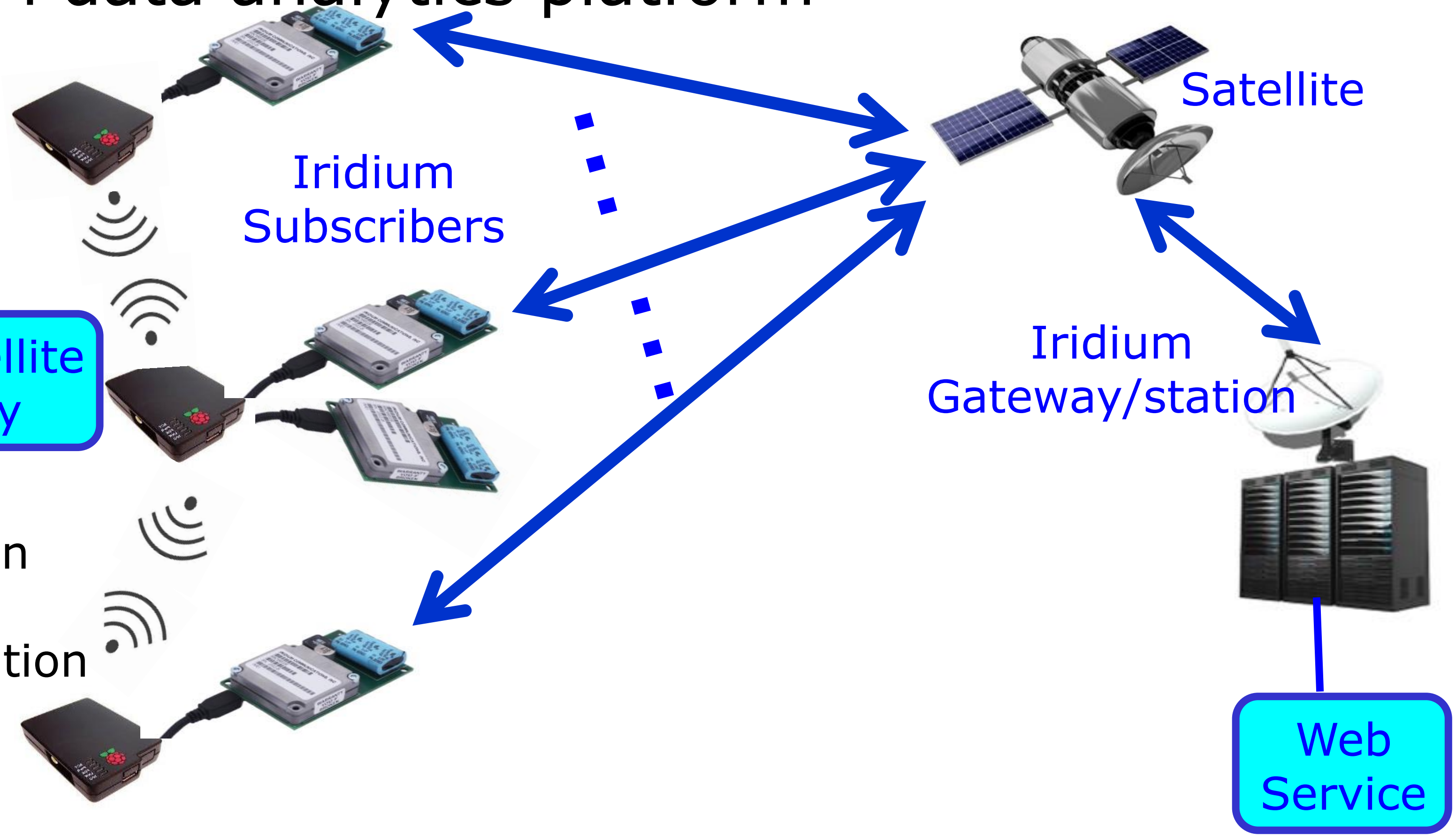
- Interface between FA and ISU is a serial connection with extended proprietary AT commands
- Interface is used to load/retrieve messages



From: <300234060698680@rockblock.rock7mobile.com>
Date: 2013/8/12
Subject: Message 4 from RockBLOCK 300234060698680
To: abraham.martin@cl.cam.ac.uk
IMEI: 300234060698680
MOMSN: 4
Transmit Time: 13-08-12 14:43:08 UTC
Iridium Latitude: 52.2577
Iridium Longitude: 0.1172
Iridium CEP: 11
Data: 7465737431
--
test1

RasPi Satellite Gateway

- Build stream processing paradigm
- RasPi data analytics platform



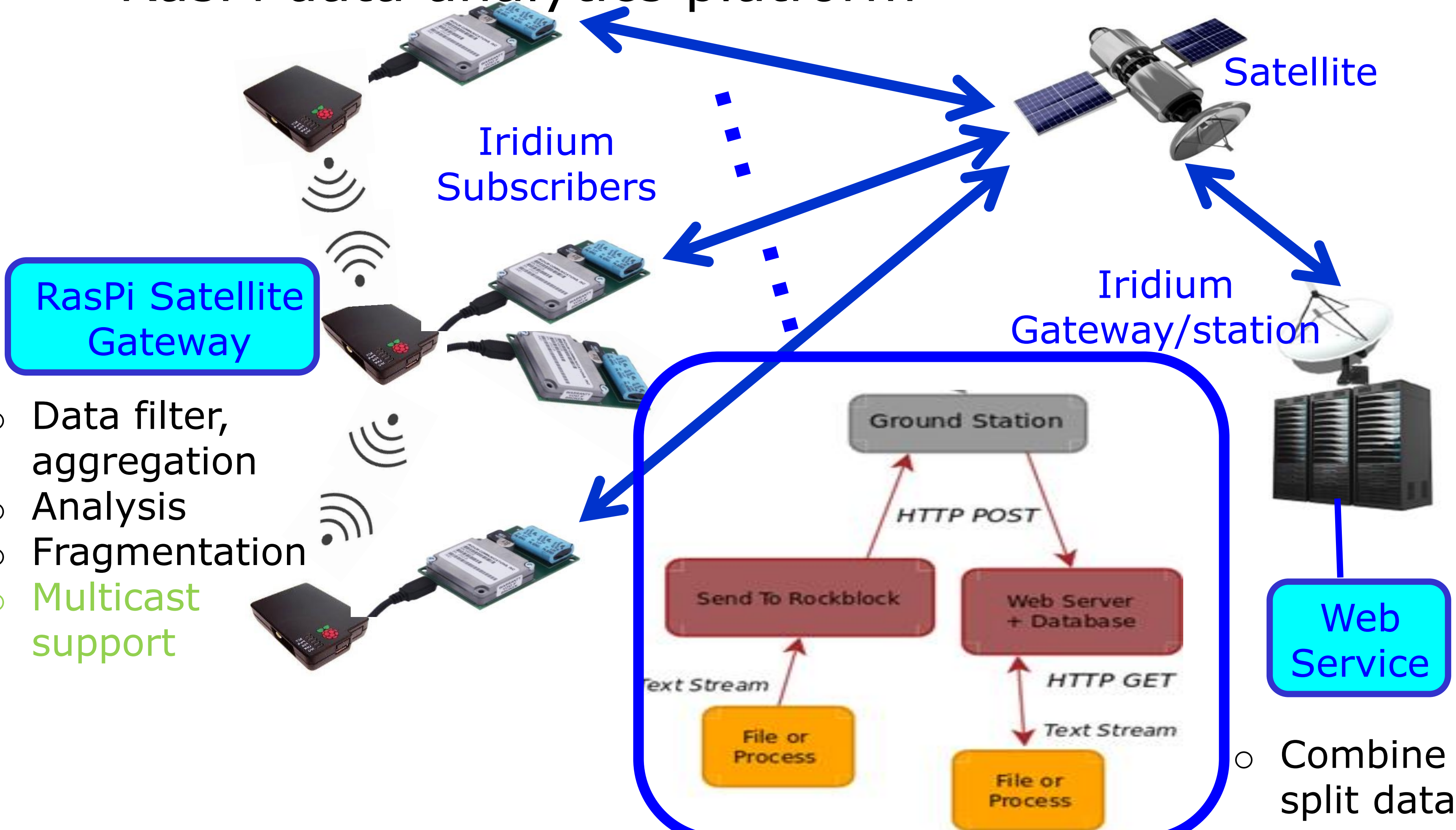
RasPi Satellite Gateway

- Data filter, aggregation
- Analysis
- Fragmentation
- Multicast support

- Combine split data

RasPi Satellite Gateway

- Build stream processing paradigm
- RasPi data analytics platform

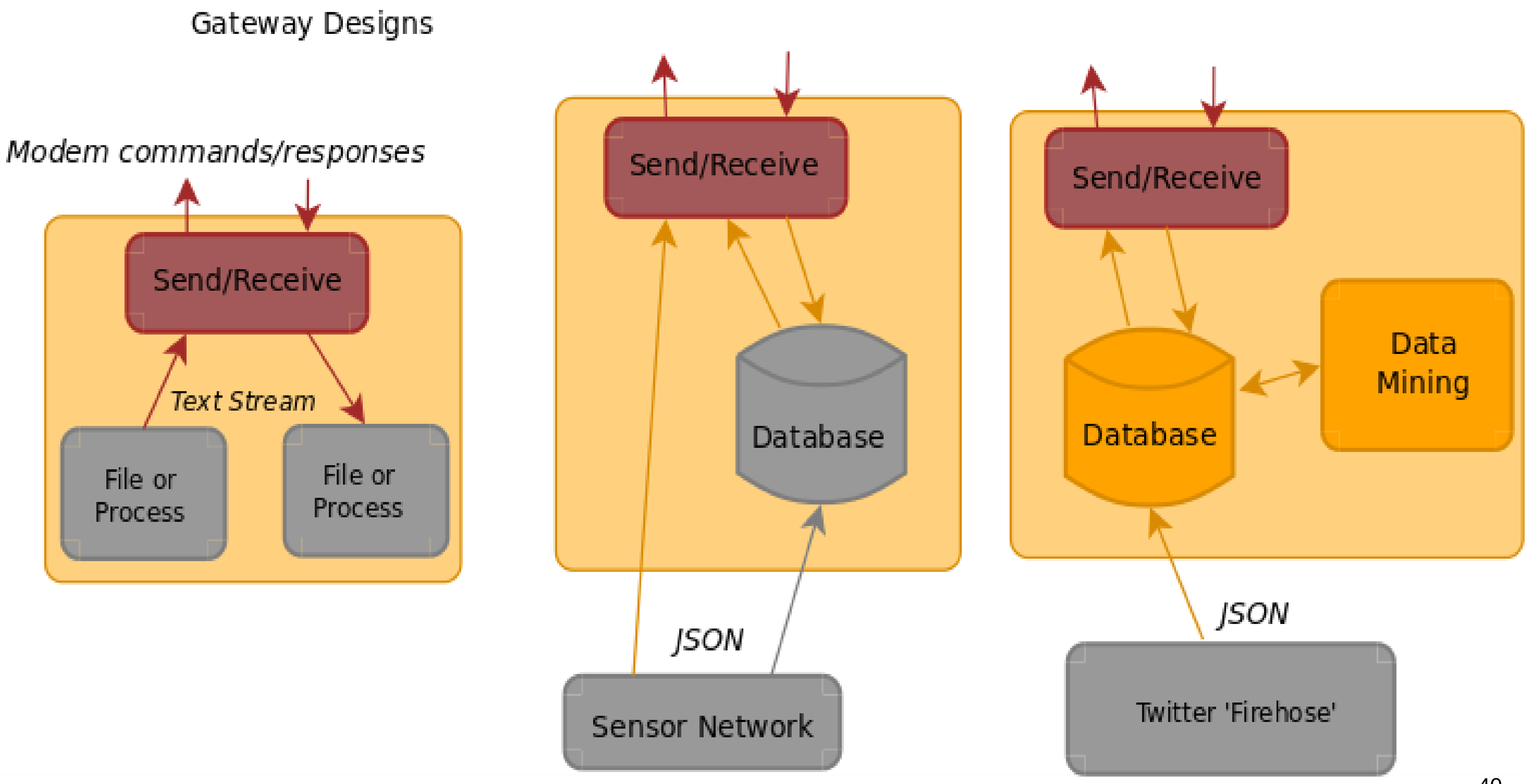


- Data filter, aggregation
- Analysis
- Fragmentation
- Multicast support

- Combine split data

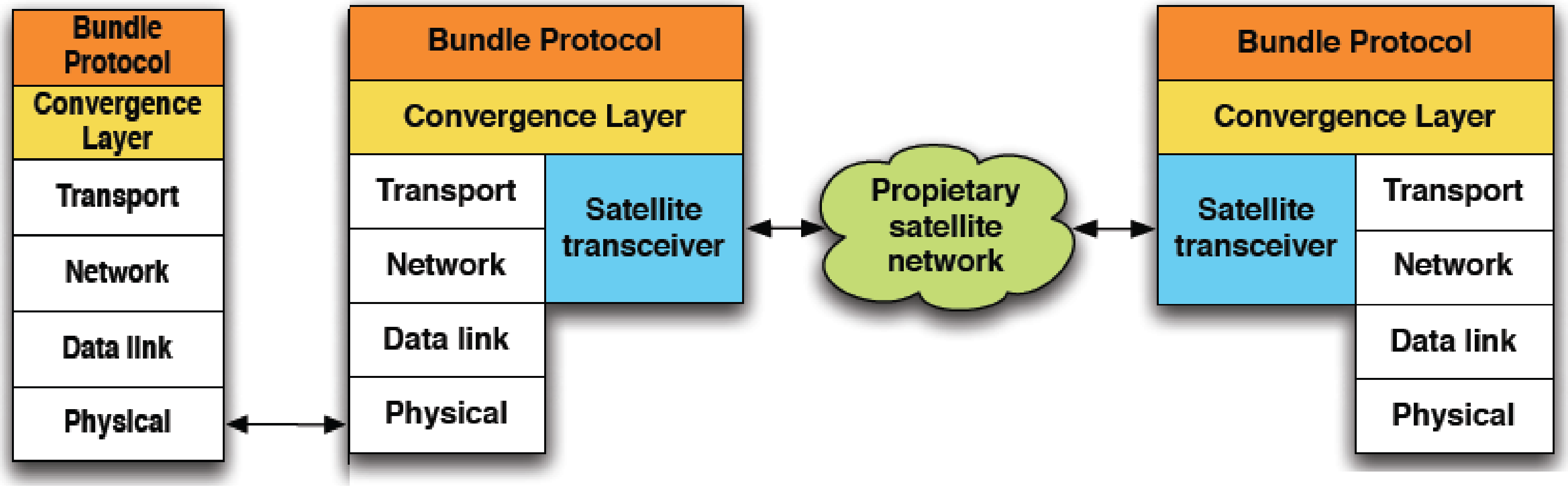
Extension to JSON Interface

- Text and JSON (converted from RFID/Twitter..)

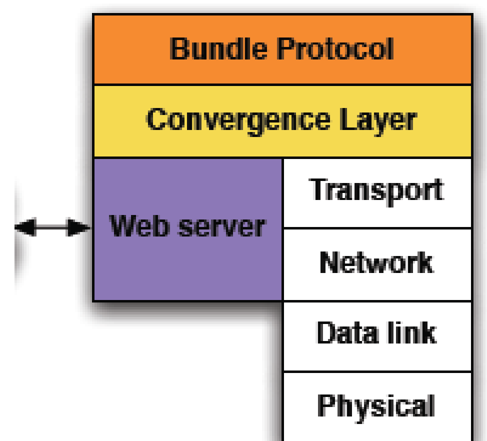


Communication Protocol

- Protocol for communication between devices with satellite transceiver



- Rockblock provides Web Service Interface
- also Email Interface



Data Compression

- Message to Iridium network < 340 bytes
- Received message < 270 bytes every 10 seconds
- DTN2 also ION
- Additional compression and fragmentation protocols are needed that are not included in the default stack of communication

Future:

- Raspberry Pi has ability of data processing
  - Cluster of Raspberry Pis for MapReduce
  - Data analysis within Raspberry Pi

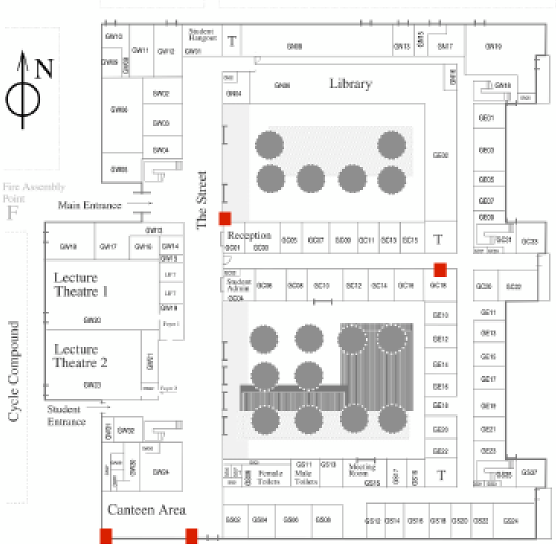
Pilot Study in Computer Laboratory

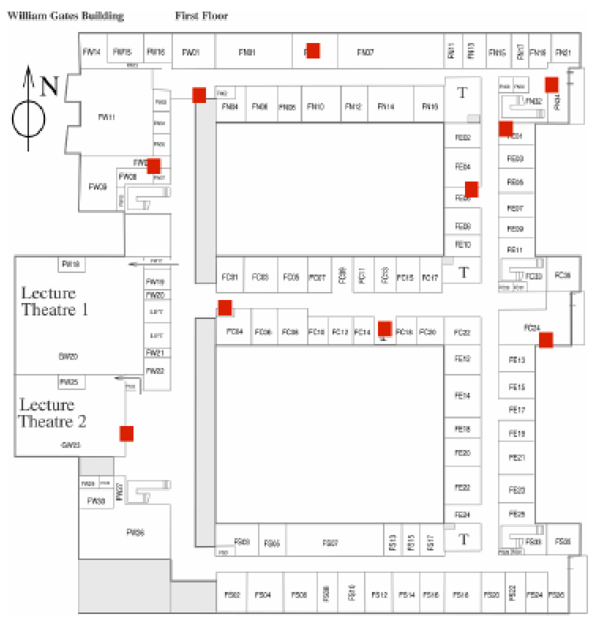
- 15 RasPi OpenBeacon Readers around Computer Laboratory
- 30 participants (4 groups)
- 3 days of data collection

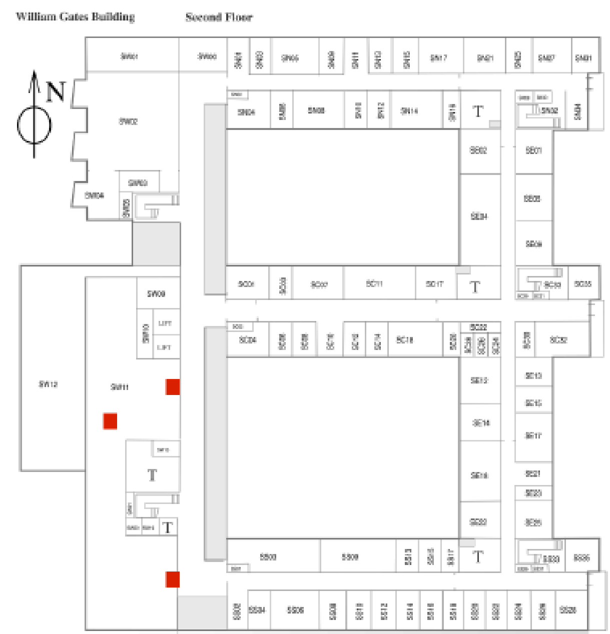
A participant wearing three RFID tags

Setting RasPiNET on 3 Floors

- Use of Data Mule approach for Data Collection
- Satellite Communication for sending statistics and changing sensing rate

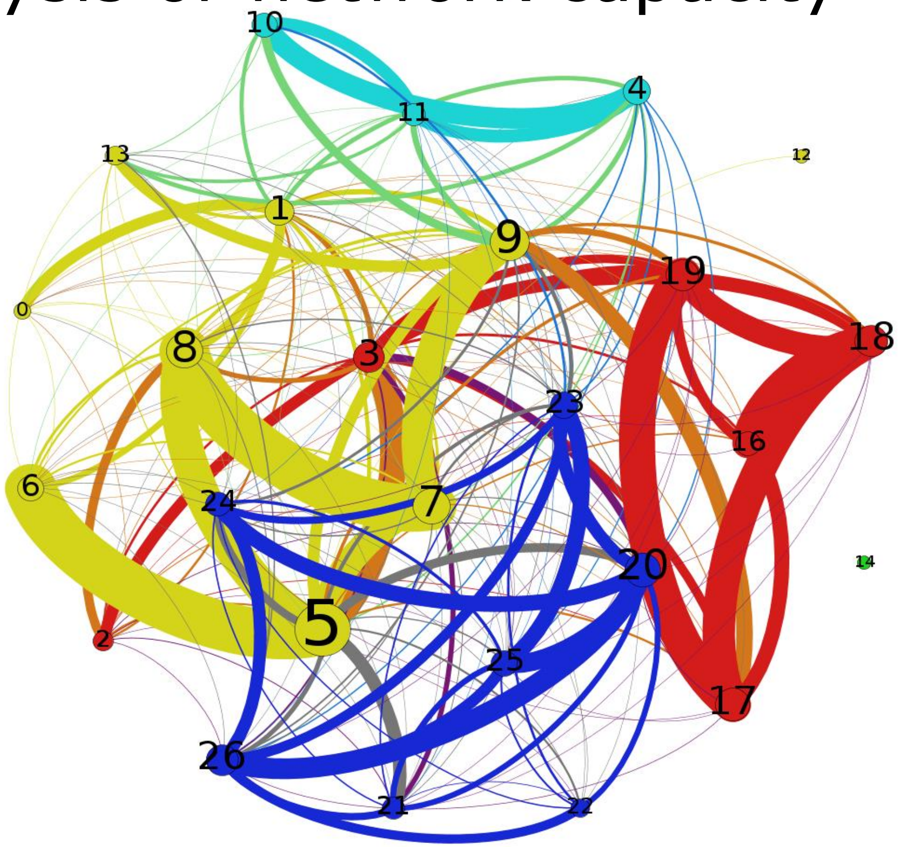






Post Data Analysis on Pilot Study

- Community Detection (4 groups and bridging nodes can be identified)
- No in-depth traffic analysis or network capacity evaluation yet
- One simulator based Simulator (w and w/o satellite connectivity)



Potential Applications?

- Form disaster recovery networks: place RasPi together with triage by first rescue responders
  - Sensing surroundings, victim's condition
  - Storing and forwarding sensed data by smart phones
- Raspberry Pi at every house: turn on for forming a local network in disaster and emergency case

Tourism Support

- Develop Research Project as a Tourism Support
  - Messaging service within Village + to/from home country
  - Visualisation of experiments

  - Use of directional antenna for P2P WiFi or Bluetooth in the village that can demonstrate messaging service or chat between the different locations within the village without Internet access plus satellite based connection to home

  - Bulk messages gathered by RasPiNET can be transmitted to Internet once a day if the bulk data in USB stick can be carried by the car daily base to the town where the Internet access exists (e.g. 200km away)

Digital Epidemiology with RasPiNet

- Rhythm and Randomness in Human Contact: http://arxiv.org/abs/1009.3980, 2010
- On Joint Diagonalisation for Dynamic Network Analysis: http://arxiv.org/abs/1110.1198, 2011
- EpiMap: Towards Quantifying Contact Networks for Understanding Epidemiology in Developing Countries: Elsevier Ad Hoc Networks Journal: Special Issue on Wireless Technology for Humanitarian Relief, 2014

RasPiNET: Decentralised Network for Data Collection and Communication with Raspberry Pi

eiko.yoneki@cl.cam.ac.uk
http://www.cl.cam.ac.uk/~ey204