Once online, our every movement, every click, sent or received email, our every activity produces a vast amount of invisible traces. These traces, once collected, put together and analysed, can reveal our behavioral patterns, location, contacts, habits and most intimate interests. They often reveal much more than we feel

comfortable sharing.

Most of those traces are hidden in **metadata**, i.e. tiny pieces of information stored in IP packets, headers of your emails or files that you are creating.

THERE IS AN ONGOING DEBATE OVER THE SIGNIFICANCE OF METADATA. WE WANTED TO QUESTION A SOMEWHAT HERETICAL ARGUMENT THAT BULK METADATA CONTAIN SENSITIVE INFORMATION ABOUT PRIVATE LIFE OF INTERNET USERS AND CONFRONT IT WITH A RULING OPINION THAT SUCH STATEMENT IS OVERRATED. WE HAVE THEREFORE UNDERTAKEN THE FOLLOWING SOCIAL AND SCIENTIFIC EXPERIMENT USING DIFFERENT METHODOLOGIES. THE PURPOSE OF THIS RESEARCH IS TO INVESTIGATE AND CONSEQUENTLY INFORM THE SCIENTIFIC AND POPULAR AUDIENCE ABOUT THE REAL IMPORTANCE OF METADATA FOR OUR PRIVACY.

In our previous research we explained how metadata is being collected and accessed by numerous actors – government agencies, Internet service providers, Internet companies such as Google or Facebook, data dealers or producers of mobile phone applications. We explained the invisible infrastructure behind data flow, but we never had a chance to investigate what these actors can really do when they have access to a vast amount of metadata about you. This data investigation is exactly about that.

THIS STORY IS ABOUT THE POWER OF METADATA.

OUR LITTLE "BIG DATA"

On July 5, 2015, one of the World's biggest cyber weapon manufacturers and dealers – an Italian based company, Hacking Team, faced a leak of their internal email database. The twitter account of the company was compromised by an unknown individual who published an announcement of a data breach and provided links to over 400 gigabytes of data, internal emails, invoices, and source code.

In the next few days Wikileaks and Transparency Toolkit published a searchable database of Hacking Team's emails revealing details of their operation, contacts and communication with government agencies, companies and individuals around the globe as well as the functionalities of their cyber surveillance weapons.

HACKING TEAM DESIGNED A MODULAR, MULTIFUNCTIONAL AND CROSS PLATFORM SURVEILLANCE SOLUTION, RCS (REMOTE CONTROL SYSTEM). THE SOLUTION GIVES THE OPERATOR A FULL AND UNINTERRUPTED ACCESS TO AND CONTROL OVER THE INFECTED DEVICE, THE PRIVILEGES OF THE OPERATOR OF THE SOFTWARE ARE LIMITLESS, THEY CAN SEND EMAIL OR SMS AND MAKE PHONE CALLS, LISTEN IN ON THE USER'S PHONE CALLS AND READ ENCRYPTED COMMUNICATION. THE ACCESS IS NOT LIMITED EXCLUSIVELY TO THE SOFTWARE, THE OPERATOR CAN ALSO MANIPULATE WITH THE HARDWARE ON THE INFECTED DEVICE, I.E. ACTIVATE THE MICROPHONE OR THE CAMERA AND RECORD AUDIO AND VIDEO OR TAKE PHOTOS. THE SOFTWARE IS DESIGNED IN SUCH A MANNER THAT ITS OPERATION GOES UNDETECTED BY ANY ANTI-MALWARE OR ANTI-SPYWARE SCANNER, ITS TRAFFIC IS WELL BLENDED IN WITH THE USER'S LEGITIMATE INTERNET TRAFFIC.

WE WERE GIVEN THIS PILE OF DATA AND SOON WE REALIZED THERE WAS ANOTHER GEM HIDDEN IN IT. WE WERE ABLE TO EXTRACT A SUBSTANTIAL AMOUNT OF METADATA – HEADERS FROM HUNDREDS OF THOUSANDS EMAILS FROM THEIR DATABASE. WE GOT OUR OWN LITTLE PORTION OF BIG DATA AND THAT IS WHERE

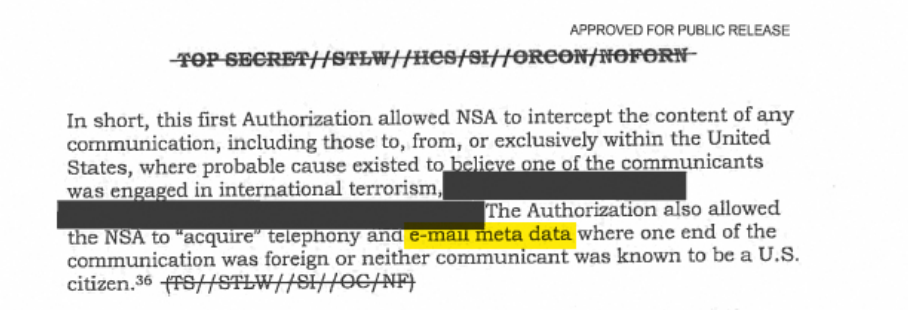
OUR RESEARCH BEGAN.

DO IT YOURSELF METADATA INVESTIGATION

The concept behind data-mining and analysis operation performed by the government agencies around the world is that metadata can be analysed to reveal connections between people, and these links can generate significant investigative leads.

This is not exclusively done by government agencies, our metadata is constantly collected and examined by major Internet companies such as Google and Facebook, but for the purpose of profiling of users and transforming our behavior into profit, which reaches tens of billions US dollars per annum.

Thanks to Edward Snowden's revelations in June 2013 we got insight into the NSA Stellar Wind, Boundless Informant, PRISM and XKeyscore programs. One of the scopes of those programs was collecting and analysing large amount of email metadata. Analysis involve operations such as contact chaining, building a network graph that models the communication (e-mail, telephony, etc.) patterns of targeted entities (people, organisations, etc) and their associates from the communications sent or received by the targets.



APPROVED FOR PUBLIC RELEASE

~~TOP SECRET//STLW//HCS/SI//ORCON/NOFORN~~

In short, this first Authorization allowed NSA to intercept the content of any communication, including those to, from, or exclusively within the United States, where probable cause existed to believe one of the communicants was engaged in international terrorism, [redacted] The Authorization also allowed the NSA to "acquire" telephony and e-mail meta data where one end of the communication was foreign or neither communicant was known to be a U.S. citizen.[36] ~~(TS//STLW//SI//OC/NF)~~

NSA IG-DRAFT REPORT

To make a pint of just how intrusive metadata analysis can be, we used this substantial amount of metadata we were able to extract from the HT's published email database, along with publicly available knowledge and a number of free or trial versions of tools available online, to conduct our own investigation.

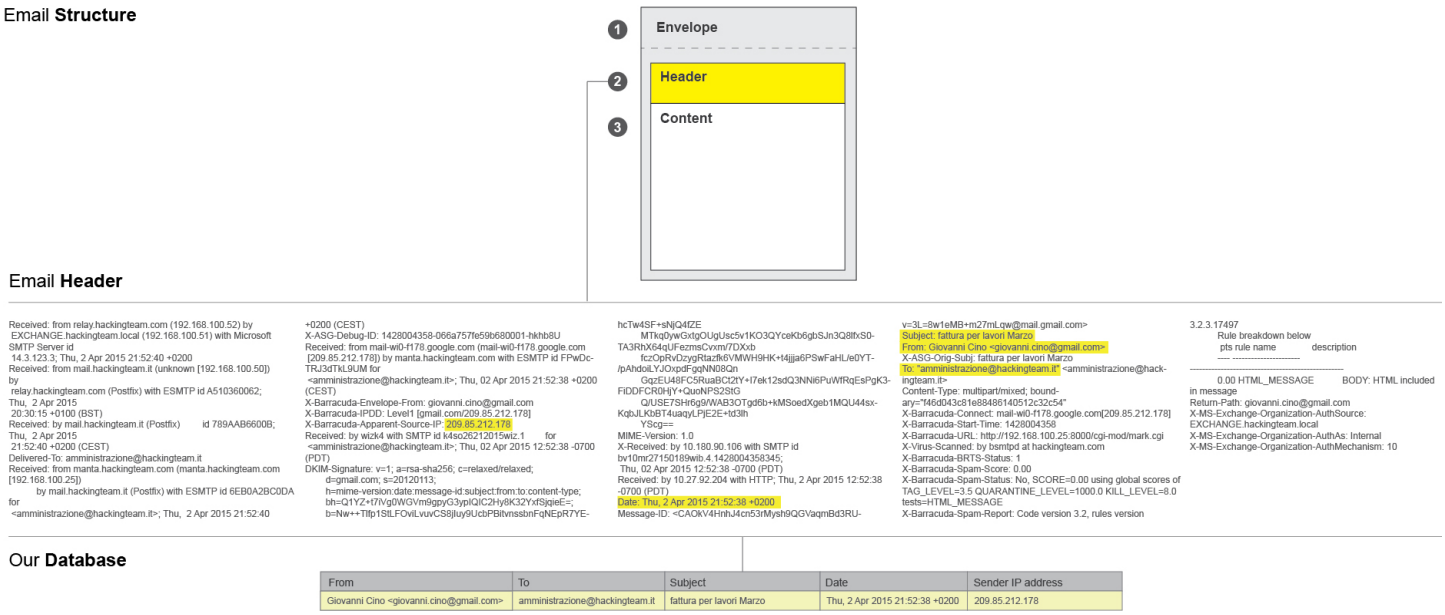
"METADATA IS EXTRAORDINARILY INTRUSIVE. AS AN ANALYST, I WOULD PREFER TO BE LOOKING AT METADATA THAN LOOKING AT CONTENT, BECAUSE IT'S QUICKER AND EASIER, AND IT DOESN'T LIE."
EDWARD SNOWDEN

IN SOME KIND OF REVERSE ENGINEERING PROCESS WE EXPLORED THE POSSIBILITY OF USING THEIR OWN METHODOLOGY FOR AN INDEPENDENT DATA INVESTIGATION OF THE HACKING TEAM, ONE OF THE "CORPORATE ENEMIES OF THE INTERNET".

INVESTIGATION PROCESS

EMAIL METADATA : BUILDING BLOCKS FOR OUR INVESTIGATION

Let's begin with a short explanation of our little treasure – email headers. Every email consists of three components: the envelope, the header, and the body of the message. The envelope is a part of the internal process by which an email is routed, the body is the actual content of the message and the header, the third component of an email, is the point of interest of our research.



Headers identify particular routing information of the message, including the sender, recipient, date and subject, sending and receiving time stamps. In some cases email headers also contain the IP address of the sender and information on the route an email takes as it is transferred from one computer to another.

	Subject	Date Sent	From(address)	To(address)	To(display)
2	Fattura HT S.R.L. n. 283158/01	12/31/2014 20:28	d.vincenzetti@hackingteam.com	fatturazioneclientige@ge.com	GE Capital Services S.r.l.
3	Fattura HT S.R.L. n. 283106/01	12/31/2014 20:28	d.vincenzetti@hackingteam.com	fatturazioneclientige@ge.com	GE Capital Services S.r.l.
4	Fattura HT S.R.L. n. 283042/01	12/31/2014 20:27	d.vincenzetti@hackingteam.com	fatturazioneclientige@ge.com	GE Capital Services S.r.l.
5	Touching Base (John Hall)	12/31/2014 18:12	d.vincenzetti@hackingteam.com	/O=HACKINGTEAM/OU=EXCHANGE ADMIN	David Vincenzetti
6	Touching Base (John Hall)	12/31/2014 18:12	d.vincenzetti@hackingteam.com	/O=HACKINGTEAM/OU=EXCHANGE ADMIN	Giancarlo Russo
7	I: Touching Base (John Hall)	12/31/2014 17:38	d.vincenzetti@hackingteam.com	/O=HACKINGTEAM/OU=EXCHANGE ADMIN	Giancarlo Russo
8	...FromSaudiArabia.	12/31/2014 14:40	d.vincenzetti@hackingteam.com	ship1208@gmail.com	ship1208@gmail.com
9	...FromSaudiArabia.	12/31/2014 13:42	d.vincenzetti@hackingteam.com	/O=HACKINGTEAM/OU=EXCHANGEADMIN	DanieleMilan
10	TTR	12/31/2014 13:24	d.vincenzetti@hackingteam.com	nupt@dhag.com.vn	Nu Pham
11	TTR	12/31/2014 13:20	d.vincenzetti@hackingteam.com	/O=HACKINGTEAM/OU=EXCHANGE ADMIN	Marco Bettini
12	HT Invoice	12/31/2014 13:20	d.vincenzetti@hackingteam.com	/O=HACKINGTEAM/OU=EXCHANGE ADMIN	Alex Velasco
13	TTR	12/31/2014 13:17	d.vincenzetti@hackingteam.com	g.russo@hackingteam.it;/O=HACKINGTEA	Giancarlo Russo;Simone
14	TTR	12/31/2014 13:14	d.vincenzetti@hackingteam.com	nupt@dhag.com.vn	Nu Pham

TOOLS : EXPORT FROM MS OUTLOOK CODE TWO OUTLOOK EXPORT > DATA PROCESSING MS EXCEL

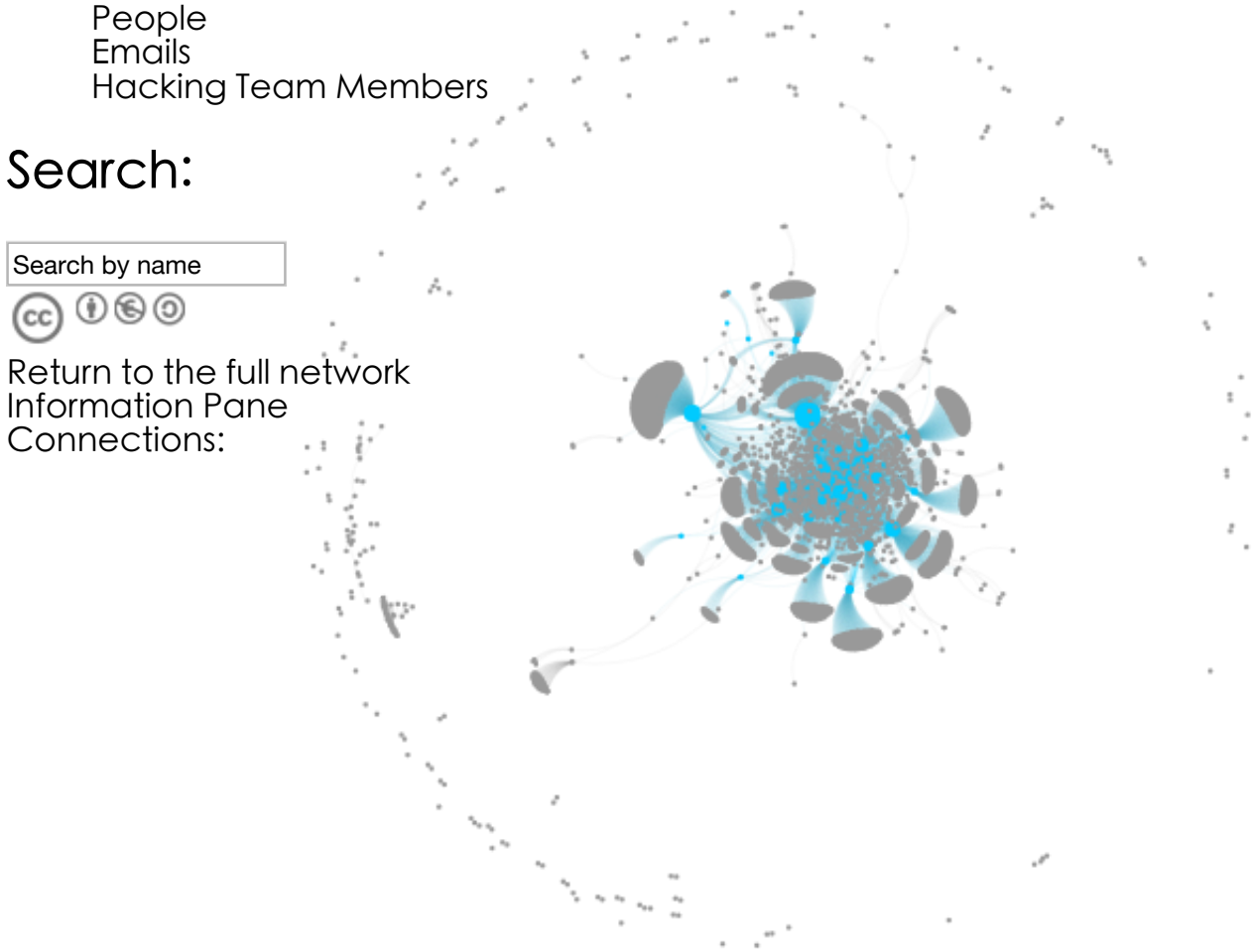
After extracting data from around 60 accounts and hundreds of thousands emails of Hacking Team employees, we got a database we could work with.

NEEDLES IN A HAYSTACK

The first step we took in exploring this pile of data was to perform a Social Network Analysis, a strategy for investigating social structures based on network and graph theories. It characterises networked structures in terms of nodes (individual actors, people) and ties or edges (relationships or interactions) that connect them. In our case, the network graph represents an analysis of all email headers exchanged between Hacking Team employees and their contacts between 2012 and 2015. Even at this very begining of the investigation we were able to detect the main internal and external actors and ties, more precisely by the amount of exchanged emails between them.

By selecting the individual nodes, we are able to explore their individual social ties and contacts.

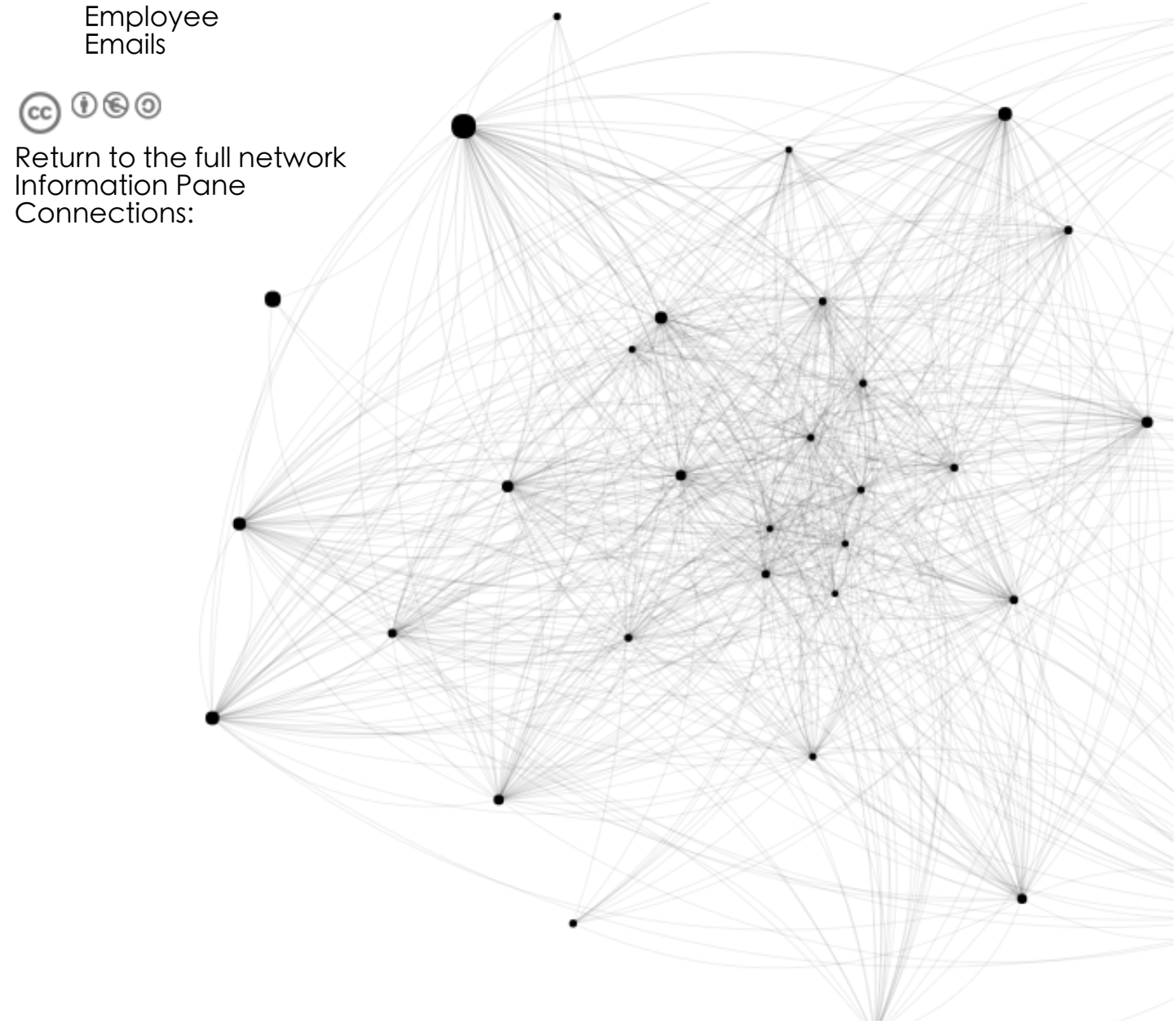
SOCIAL NETWORK ANALYSIS OF HACKING TEAM EMAIL DATABASE (PERIOD 2013-2015)



TOOLS : SOCIAL NETWORK ANALYSIS WITH GEPHY EXPORTED WITH SIGMA.JS BY OXFORD INTERNET INSTITUTE

By filtering out the nodes with less than a 100 exchanged emails, we isolate the internal Hacking Team communication and get a closer look at their internal structure based solely on it.

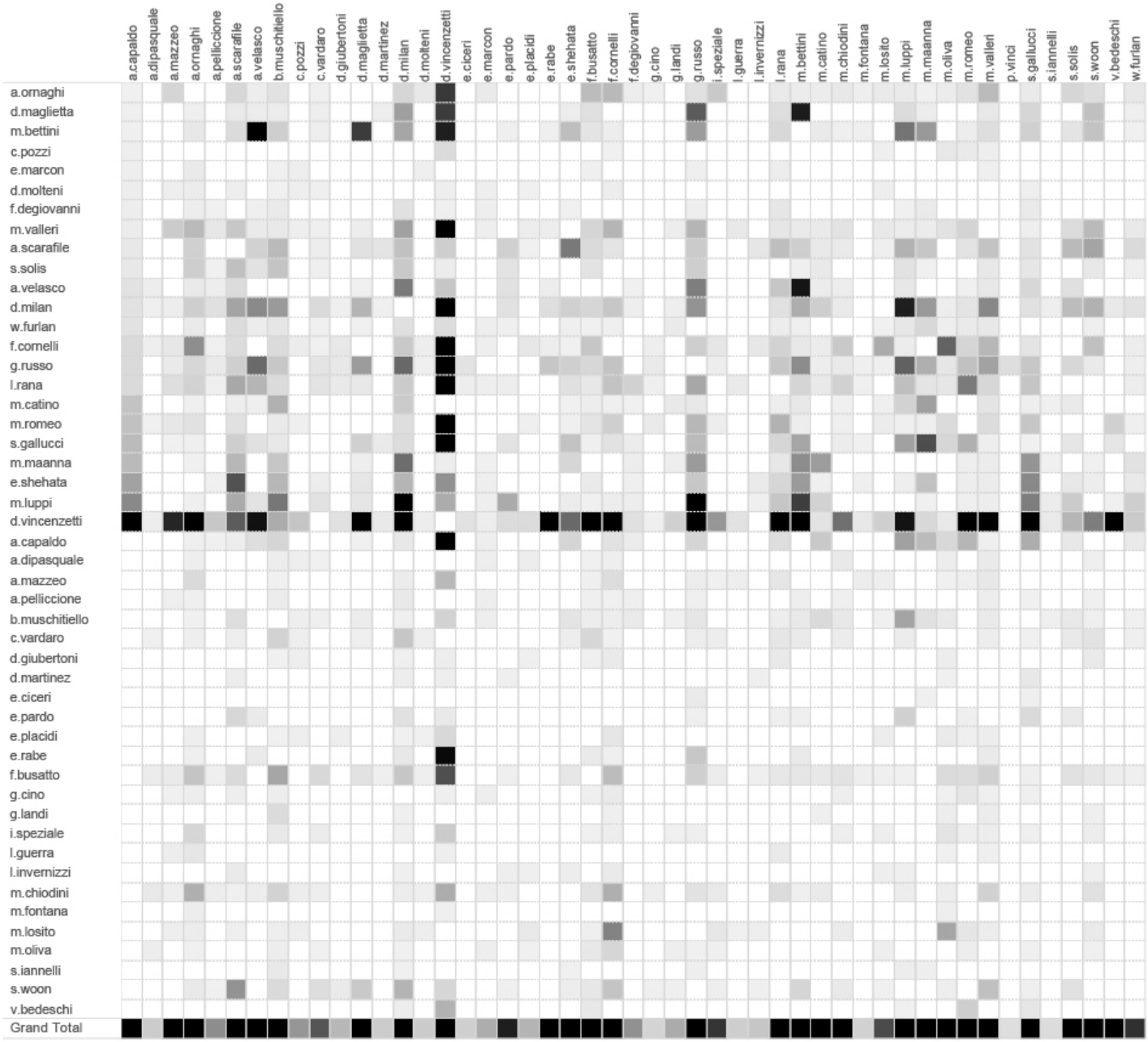
SOCIAL NETWORK ANALYSIS OF NODES WITH +100 EXCHANGED EMAILS



TOOLS : SOCIAL NETWORK ANALYSIS WITH GEPHY EXPORTED WITH SIGMA.JS BY OXFORD INTERNET INSTITUTE

Giving that this was somewhat a learning process of our own, while exploring the metadata we came to notice that our understanding of data and leads we got from it highly depended on the type of visualisation we applied to the data set. Sometimes ties between different actors were more successfully, more clearly revealed by using different visualisations. Like in this example, where we see the same data set as the one presented above, but this time in the form of a heat map.

HEAT-MAP OF INTERNAL COMMUNICATION

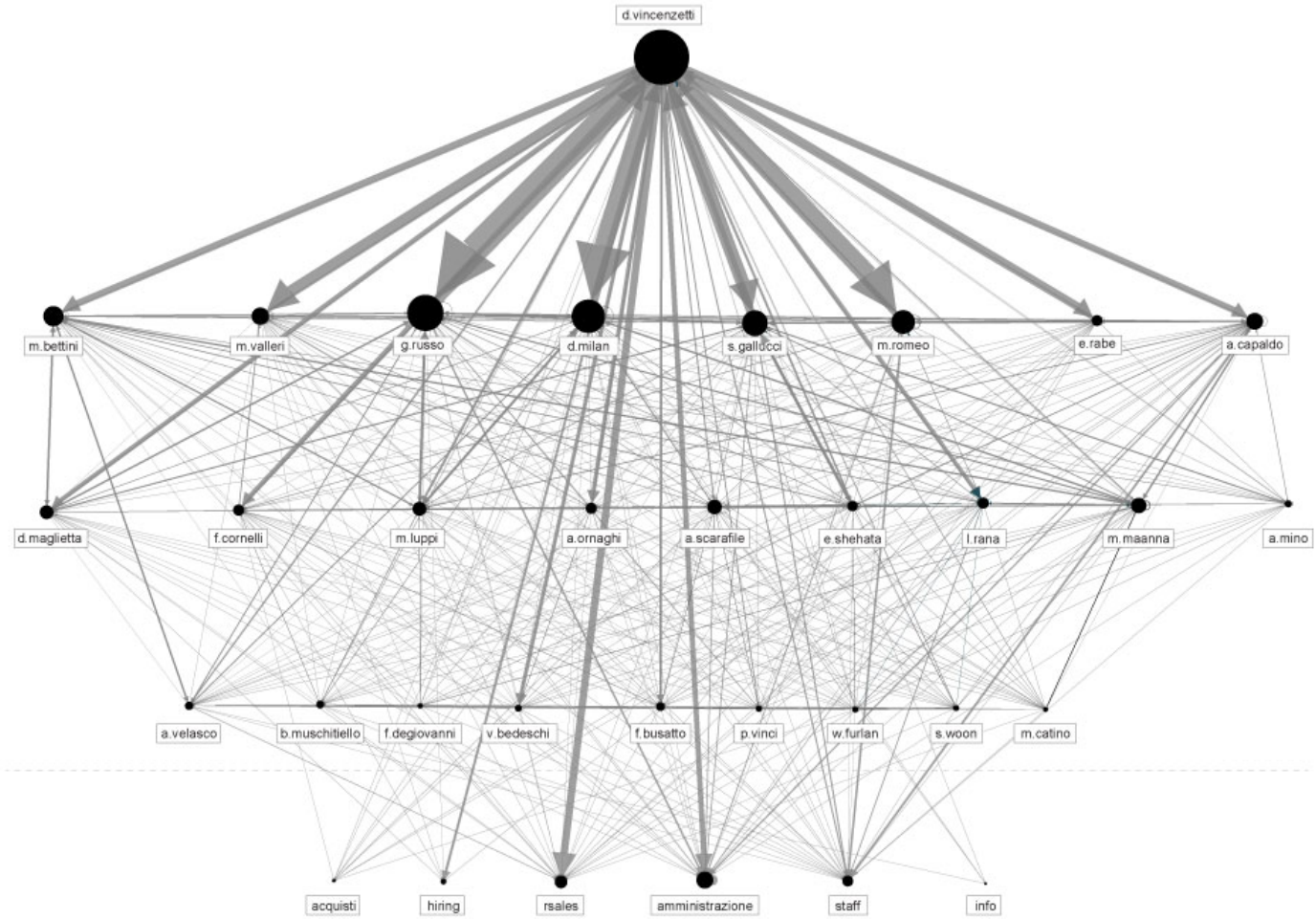


By spotting the darker squares we can explore individual ties between different employees within organisation. D. Vincenzetti is clearly the main actor in this graph, but we can also notice a few other strong relations across this heatmap, that can help us get a better insight into their organisational structure.

Finally, using the insights from both visualisation methods, we are able to shape a communication chart that might represent a credible representation of the organisational structure. It probably doesn't display relations that are in accordance with what is written on their business

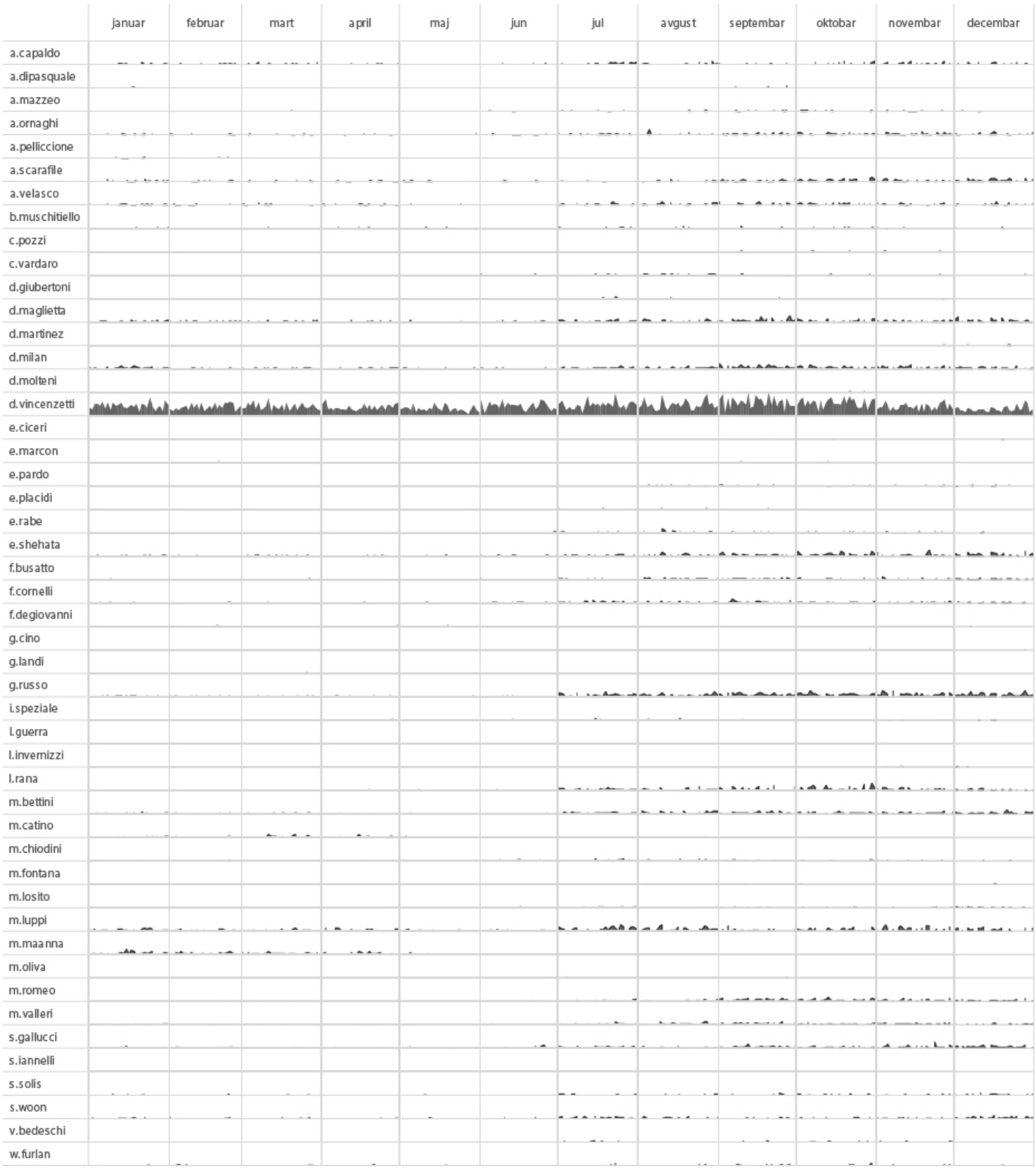
cards, but on the other hand it probably represents real relations between people within the organisation better.

POTENTIAL ORGANISATIONAL STRUCTURE BASED ON THE LEVEL AND DIRECTION OF COMMUNICATION



Setting aside the organisational structure, if we were to add another interesting piece of information retrieved from metadata – the time component, we would be able to track the activity of every individual employee in time, based on the number of sent messages from each one of them. Having done this, we created the following activity chart. With this kind of analysis you could, for example, speculate or determine which part of the year is the busiest for the organisation or, combined with other information inputs, when certain employees went on vacation or took a leave of absence.

NUMBER OF SENT EMAILS PER HT EMPLOYEE IN TIME (2014)



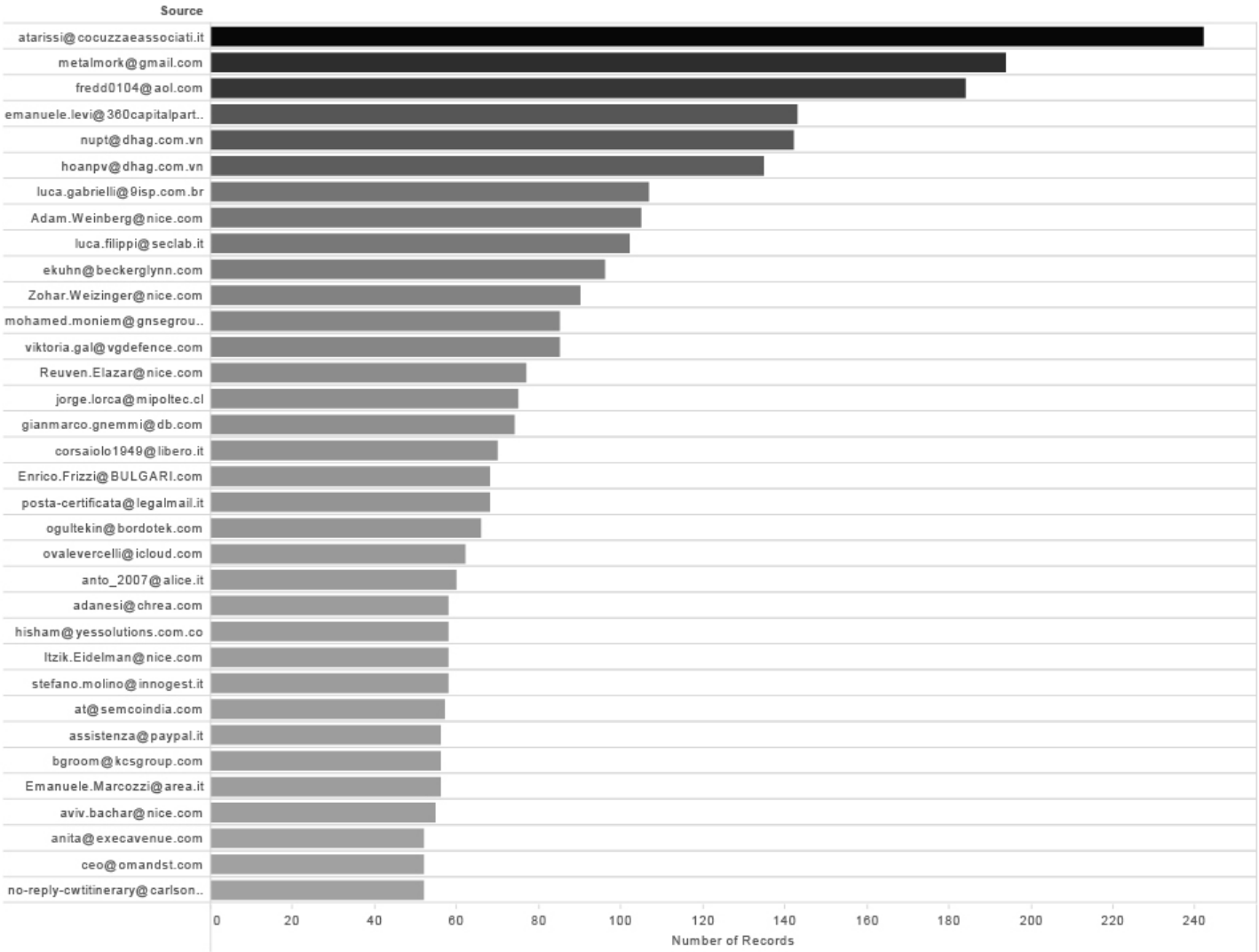
EXPLORING EXTERNAL CONTACTS

Even more interesting, or relevant for investigative data journalism and our

effort to understand the nature of the organisation that we are investigating, are probably the external contacts.
In our set of data that means around 4600 different individuals that exchanged emails with Hacking Team employees in the course of 2 years.

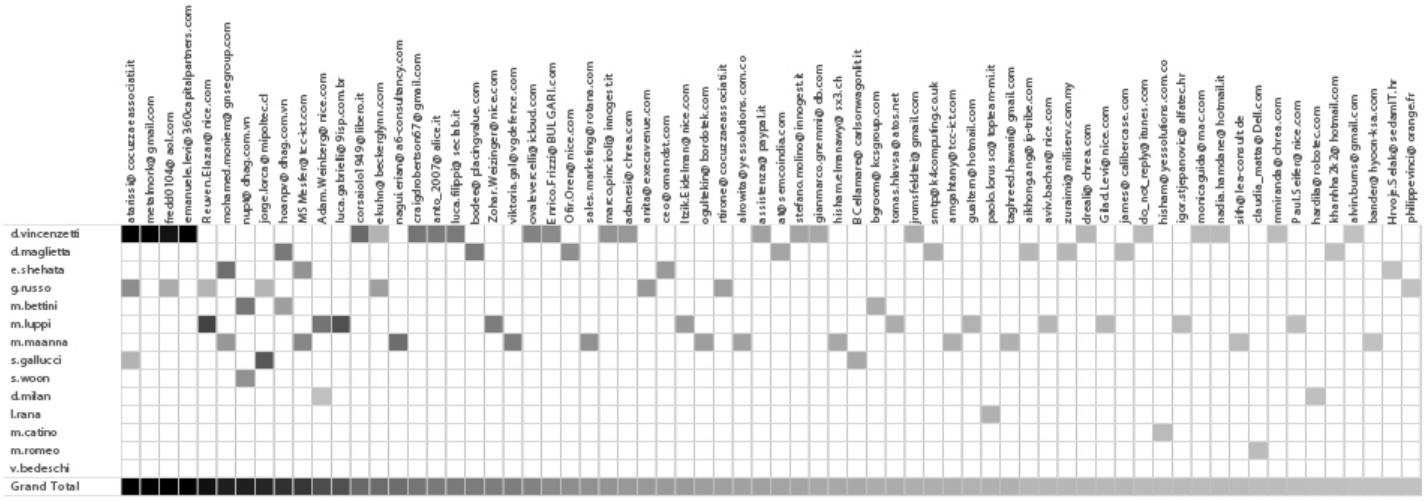
If we exclude all the @hackingteam.com adresses and rank results by the number of records we are going straight to the point. This is the list of Hacking Team contacts with more than 50 emails exchanged.

EXTERNAL CONTACTS WITH MORE THAN 50 EMAILS EXCHANGED WITH HT EMPLOYEES (2014-2015)



If we add the Hacking Team employees on the other axis, we will get information who in the team communicated with external contacts and how frequent and strong the communication was.

NUMBER OF EMAILS EXCHANGED (>30) BETWEEN HT EMPLOYEES AND EXTERNAL CONTACTS (2014-2015)

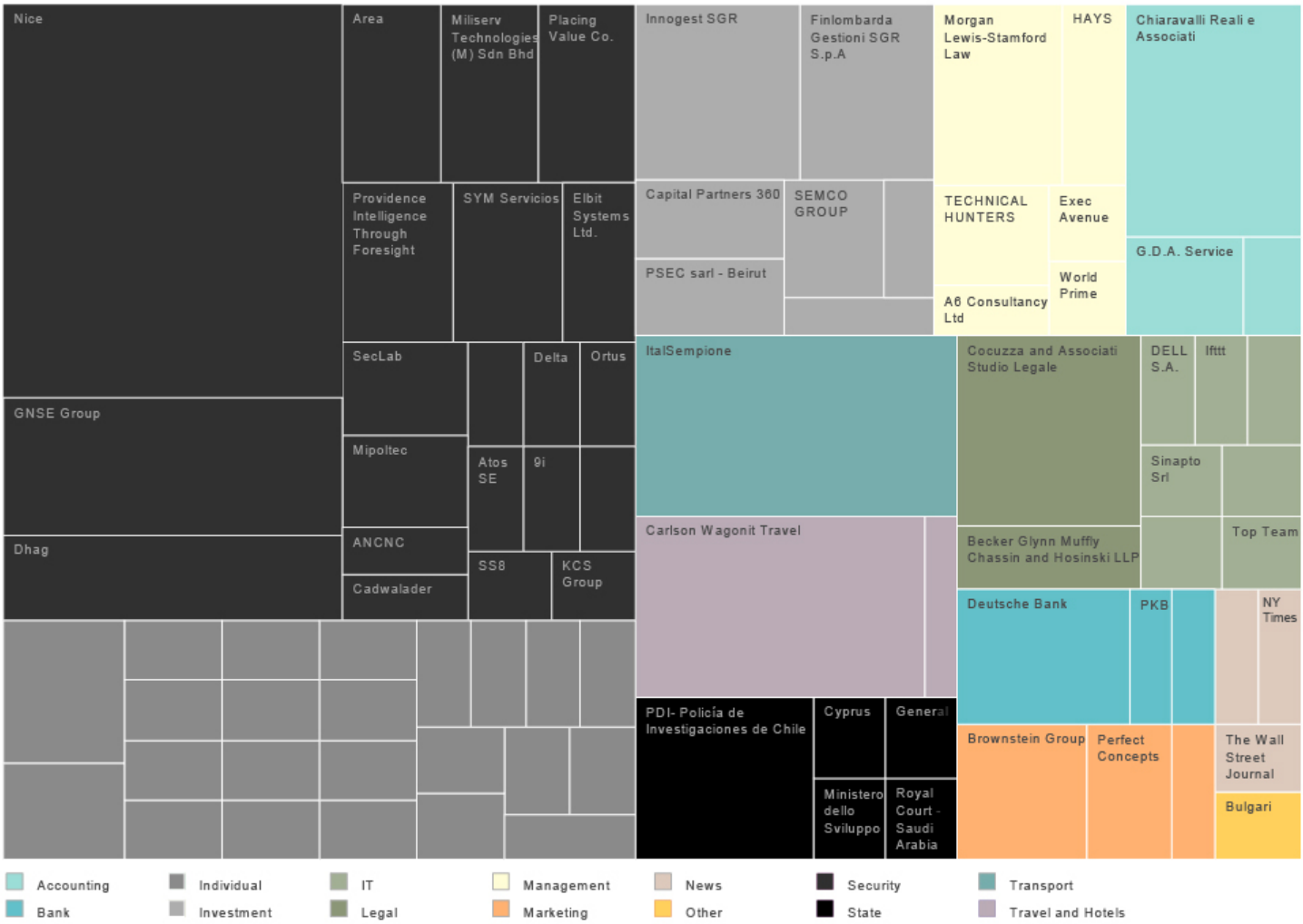


Additionally, if we add the time component, we have a complete overview of who communicated with whom and when.

TIMELINE OF INDIVIDUAL COMMUNICATION OF EXTERNAL CONTACTS AND HT EMPLOYEES (2014-2015)

We grouped the emails by domain, and after some research about the companies behind the domain names, we classified them by the type of service they officially provide.

EXTERNAL CONTACTS GROUPED BY THE DOMAIN NAME BASED ON THE D.VINCENZETTI EMAILS



According to this treemap the biggest group of organisations collaborating with HT are from the “digital security” sector, followed by individual contacts, i.e. “consultants” and venture capital companies.

We can explore the relation between selected companies and Hacking Team in time.

TIMELINE OF SELECTED COMPANIES EMAIL COMMUNICATION WITH HT EMPLOYEES (2014)

Different colours represent different people within an organisation. We can notice that, for example in the case of Nice Solutions, the biggest partner of Hacking Team according to the examined metadata, the connections between Hacking Team and Nice are really tight and there is a constant communication between the two companies including a lot of different employees on all levels (number of different colours on the graph). We can also track how different actors are taking communication in different times.

ACCORDING TO THEIR OFFICIAL STATEMENT "NICE SOLUTIONS EMPOWER ORGANISATIONS TO CAPTURE, ANALYSE, AND APPLY, IN REAL TIME, INSIGHTS FROM BOTH STRUCTURED AND UNSTRUCTURED BIG DATA. THIS DATA COMES FROM MULTIPLE SOURCES, INCLUDING PHONE CALLS, MOBILE APPS, EMAILS, CHAT, SOCIAL MEDIA, VIDEO, AND TRANSACTIONS. NICE SOLUTIONS ARE USED BY OVER 25,000 ORGANISATIONS IN MORE THAN 150 COUNTRIES, INCLUDING OVER 80 OF THE FORTUNE 100 COMPANIES". LOOKING AT THE DATA WE ANALYSED WE CAN PROBABLY CONCLUDE THAT THE SAME EXPERTISE AND TOOLS FOR COLLECTING DATA AND ANALYSIS IS SHARED AND BEING SOLD ON DIFFERENT MARKETS AND TO DIFFERENT TARGET GROUPS, HACKING TEAM FOR GOVERNMENT AGENCIES AND NICE SOLUTIONS FOR COMPANIES AND LAW ENFORCEMENT AGENCIES.

PATTERN OF LIFE

Humans are amazing pattern-recognition machines. We are constantly analysing complex sets of inputs, and making decisions based on facts we previously encountered or learned. But in recent years, we are not the only ones who analyse patterns around us, we are becoming more and more the object of analysis, mostly performed by machines and algorithms.

THE UNIQUE WAY WE INTERACT WITH THE TECHNOLOGY WE USE, THE UNIQUE SET OF CONTACTS WE HAVE OR OUR UNIQUE BEHAVIORAL PATTERNS DEFINE OUR METADATA SIGNATURE, OUR FINGERPRINT. IN THE EYES OF THE ALGORITHMIC ANALYSIS EVERY SINGLE PERSON IS UNIQUE.

Pattern-of-life analysis is a method of surveillance specifically used for documenting or understanding subject's habits. It is a computerised data collection and analysis method used to establish the subject's past behavior, determine its current behavior, and predict its future behavior. This form of analysis is generally done without your consent, and it's applied not just in the security field, but it is a core activity and business model of many of the biggest Internet companies. More commonly, pattern of life analysis is called profiling. Inputs for this analysis are in most of the cases our metadata collected from emails, IP traffic or data from mobile phones and other technology we use.

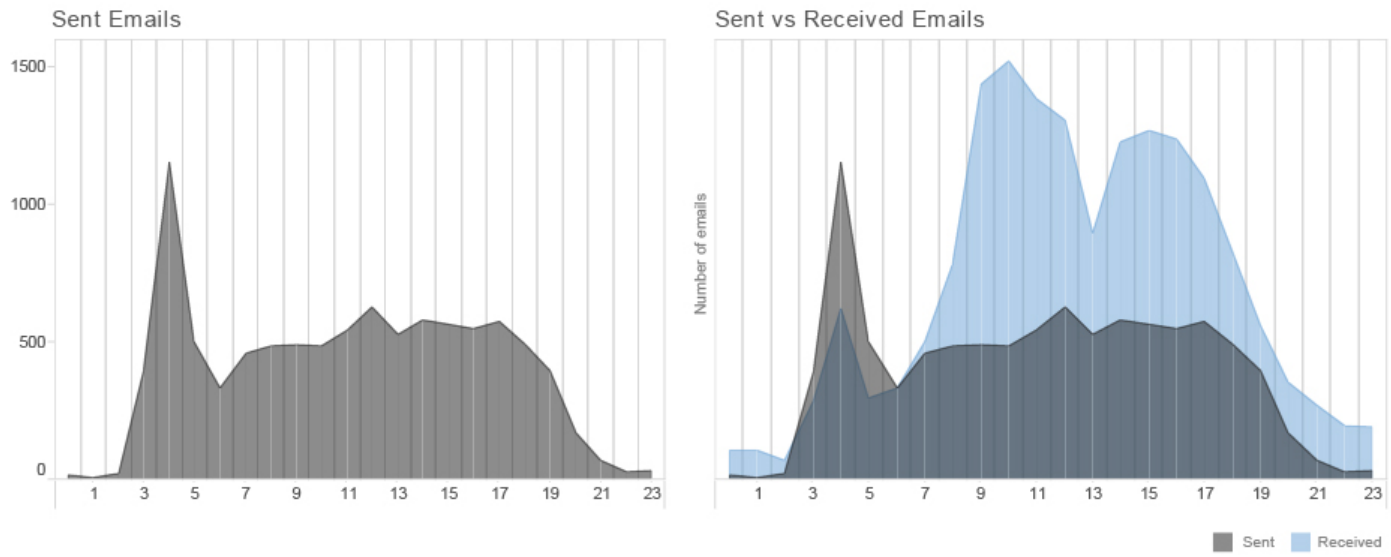
Even though we are just limited to email metadata in our research , we will try to perform pattern of life analysis on one key figure from Hacking Team and try to see what we can get.

According to the previous phases of our metadata investigation, an obvious choice of a node (person) with the biggest amount of internal and external contacts and communication is d.vincenzetti@hackingteam.com. We will call him Mr.D.

We are trying to understand 3 main things here:

What is the behavioural pattern of Mr.D ?
What are the anomalies in this pattern?
How different the behavioural pattern of Mr.D is in comparison to his social/professional circle?

PATTERN RECOGNITION : SUM OF MR.D SENT AND RECEIVED EMAILS PER HOUR DURING THE DAY (2014)

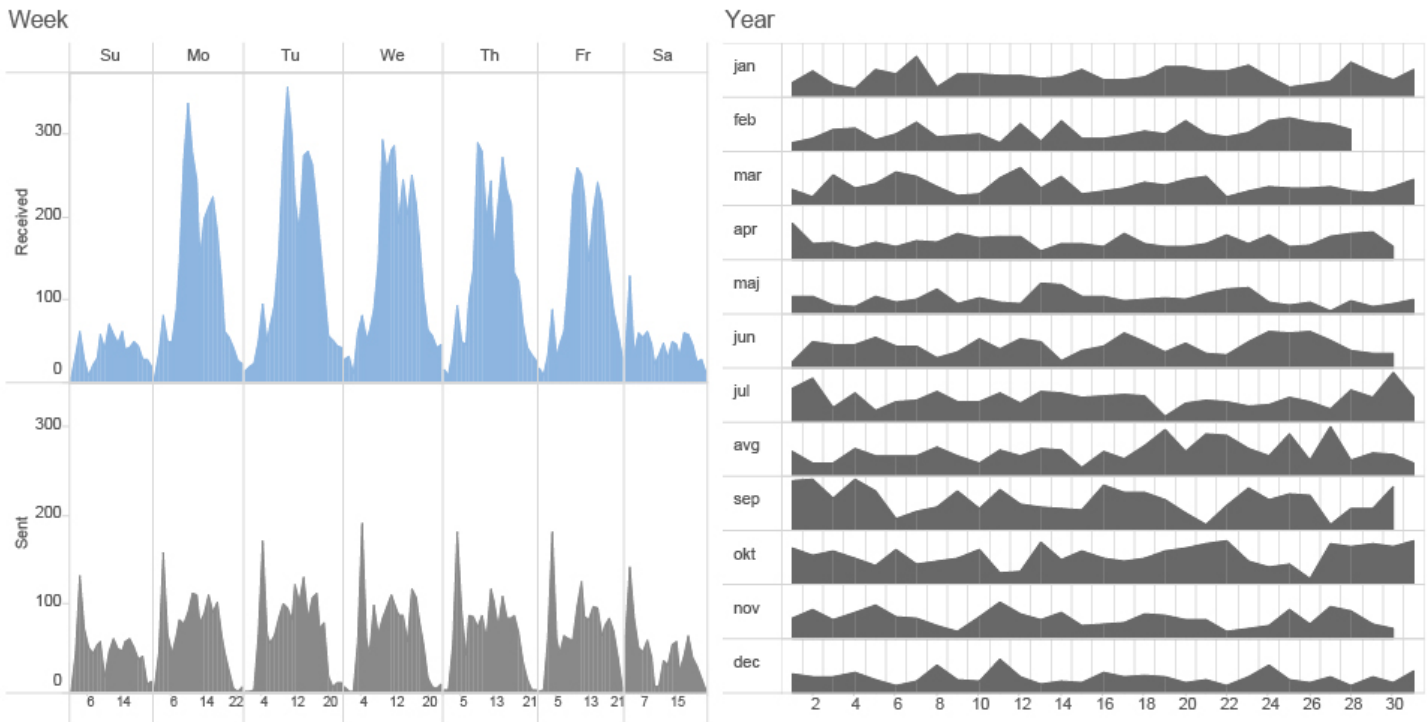


Sent emails represent the behaviour of the person that we are examining and received emails represent the overall behavioural pattern of his social or professional environment.

Mr.D is not the same as other people.

He starts his activities quite early in the morning. Almost every day around 4 a.m. is his time of concentration, the moment when he sends the biggest amount of emails during the day. If we are comparing the number of sent and received emails we can see that Mr.D has different habits than most of his contacts. His social and professional circles are most productive around 10 a.m., most of them have a lunch break around 1 p.m. and their productivity rapidly declines from 4 p.m. On the other hand, Mr.D doesn't have a big swings of productivity during the day. His peak during working hours is at noon. It looks like he doesn't have a regular lunch break and when his co-workers and external contacts start to lose concentration in the afternoon, he has another peak of activity around 5 p.m.. Additionally, Mr. D seldom sends any emails after 8 p.m.

PATTERN RECOGNITION : SUM OF MR.D SENT AND RECEIVED EMAILS PER WEEK DAYS AND MONTHS (2014)

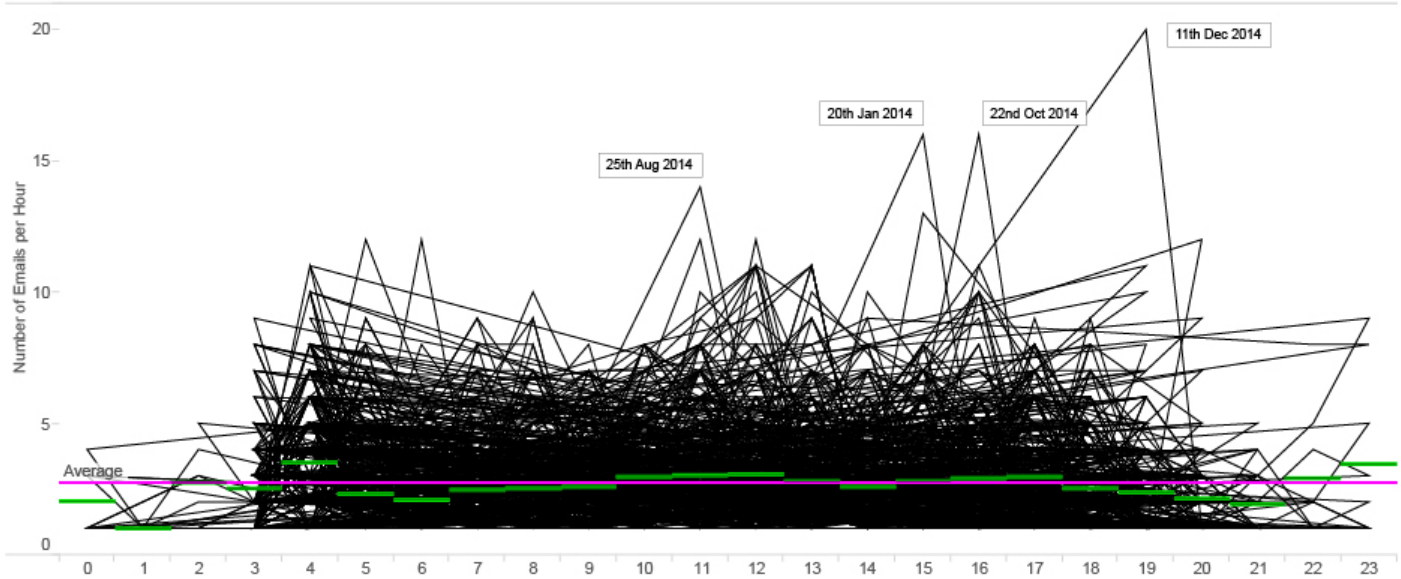


If we are analysing Mr.D's behavioral patterns on a week's scale we can find out that he is even working much more than his professional circles during the weekend as well. It looks like his only time out of emails is every

saturday during lunch time around noon.

That is Mr.D's average behavior, but what is even more important to our analysis are the anomalies in his behavior. Anomalies can point to many things. People are changing their behaviour when depressed, sick, working under pressure, when there are some deadlines or important events, when they are traveling or when they fall in love, for example.

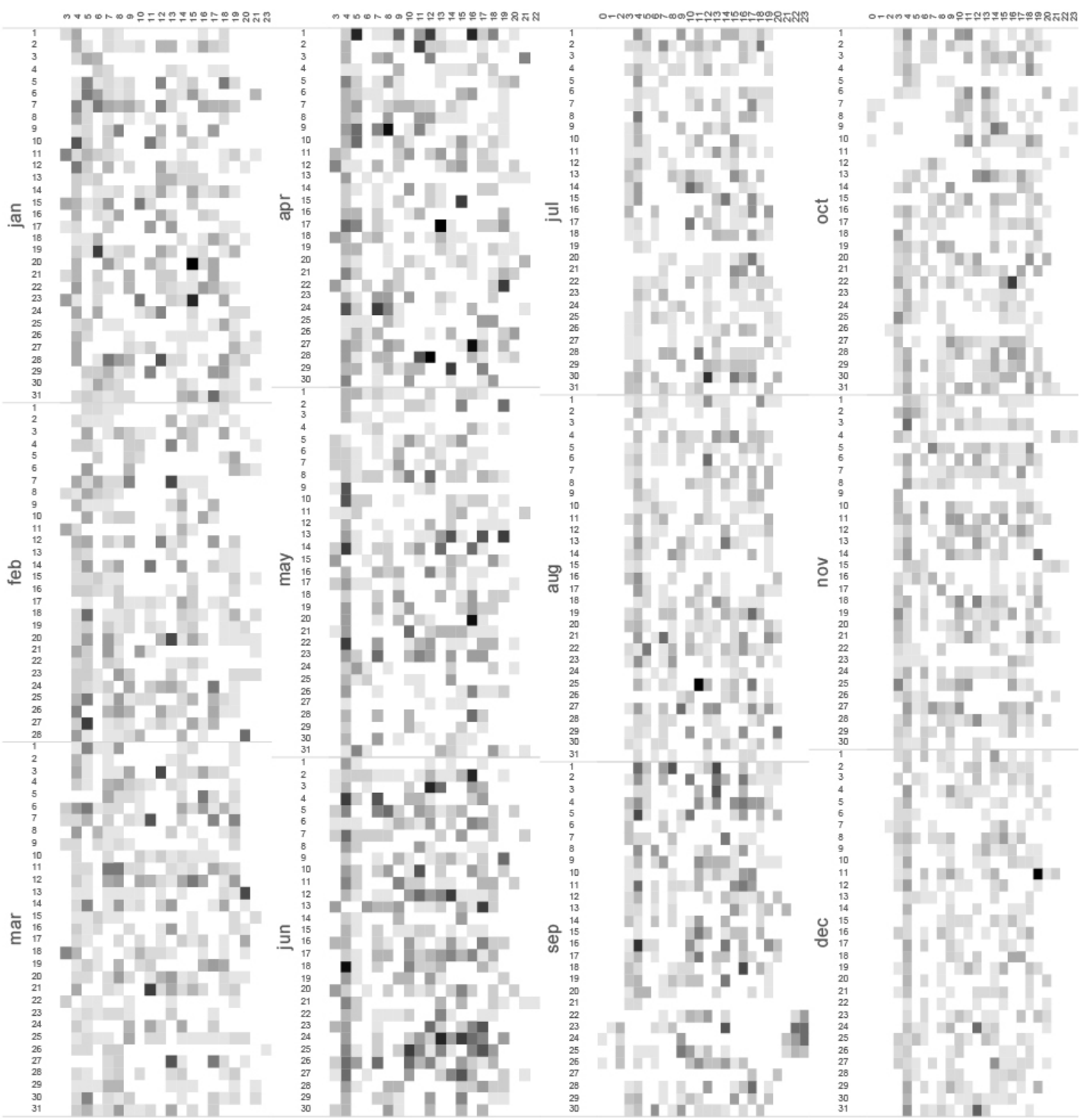
ANOMALY DETECTION : NUMBER OF MR.D SENT EMAILS PER HOUR (2014)



On this graph we can see some spikes that represent anomalies in Mr.D's productivity.
For example, on the 11th of December at 7 p.m., he sent 19 emails during one hour compared to average 2.7 emails that he usually sends.

The following heatmap is probably most effective for spotting anomalies.

PATTERN RECOGNITION AND ANOMALY DETECTION : HEAT-MAP OF MR.D SENT EMAILS PER HOUR (2014)



Looking at the heatmap, you can spot an interesting anomaly in September. On the 20th in the afternoon we see a really low level of activity, which is unusual for Mr.D, giving that we already know that his daily productive peak is around noon.

Further, on September 21st from 4 a.m, when he generally wakes up and starts work, there is no activity until the day after at 7 a.m.

In the next few days we see that the pattern is uncommon. It looks as if

Mr.D sleeps in the afternoon and works intensively during the night. On September 27th, we can see again a lack of communication and then in the following days, Mr. D's pattern is back to normal. Based on the other metadata inputs that we will explain later in our research, we found out that Mr.D was traveling to Singapore between the 21st and the 27th of September. With that we can easily conclude that a change in the time zone influences Mr.D's email pattern.

We can find another anomaly in the pattern from October 5th to October 12th, but this pattern looks a bit different, it swings in another direction. On this occasion Mr.D went to the USA and it showed us how different time zones leave different footprints in his pattern.

There is also one really interesting pattern anomaly on June the 24th and the following day (you can see darker squares and an increased level of communication). On that day, researchers from Citizens Lab published analysis "Police Story: Hacking Team's Government Surveillance Malware" exposing the functionality and architecture of Hacking Team's Remote Control System (RCS) in a never-before published detail. This report had a great media coverage, including media such as The Economist, Associated Press, Wired, VICE, International Business Times, Forbes and others. We can see how this stressful event for Hacking Team reflects on the pattern of Mr.D's emails.

MYSTERIES OF THE SUBJECT

Aside from the defined activity patterns and discovered anomalies, email subjects also reveal a very detailed overview of Mr.D's communication with other employees of Hacking Team.

TIMELINE OF EMAILS SENT FROM MR.D TO INDIVIDUAL HT EMPLOYEES IN 2014 (HOVER ON GRAPH FOR SUBJECT LINE)

TOOL : TABLEAU

Just to make a short but interesting digression: while creating this graph of email subjects we stumbled upon the moment A.Pelliccione left Hacking Team in March 2014. At this moment his communication with Mr. D stops. Based on the IP location data that we will present later, we also found out

that he moved to Malta and started communicating through a different email address – reaqta.com.

We can argue whether the email subject should be considered metadata or not. However, looking from a technical point of view, the subject is a part of the header in the same way as other types of information (From, To, Date, etc.). Basically, it's just a matter of choice of the person who is to analyse the metadata. For the intents and purposes of our research, we will consider the subject a legitimate source for metadata analysis.

TIMELINE OF SUBJECT LINES (2014)

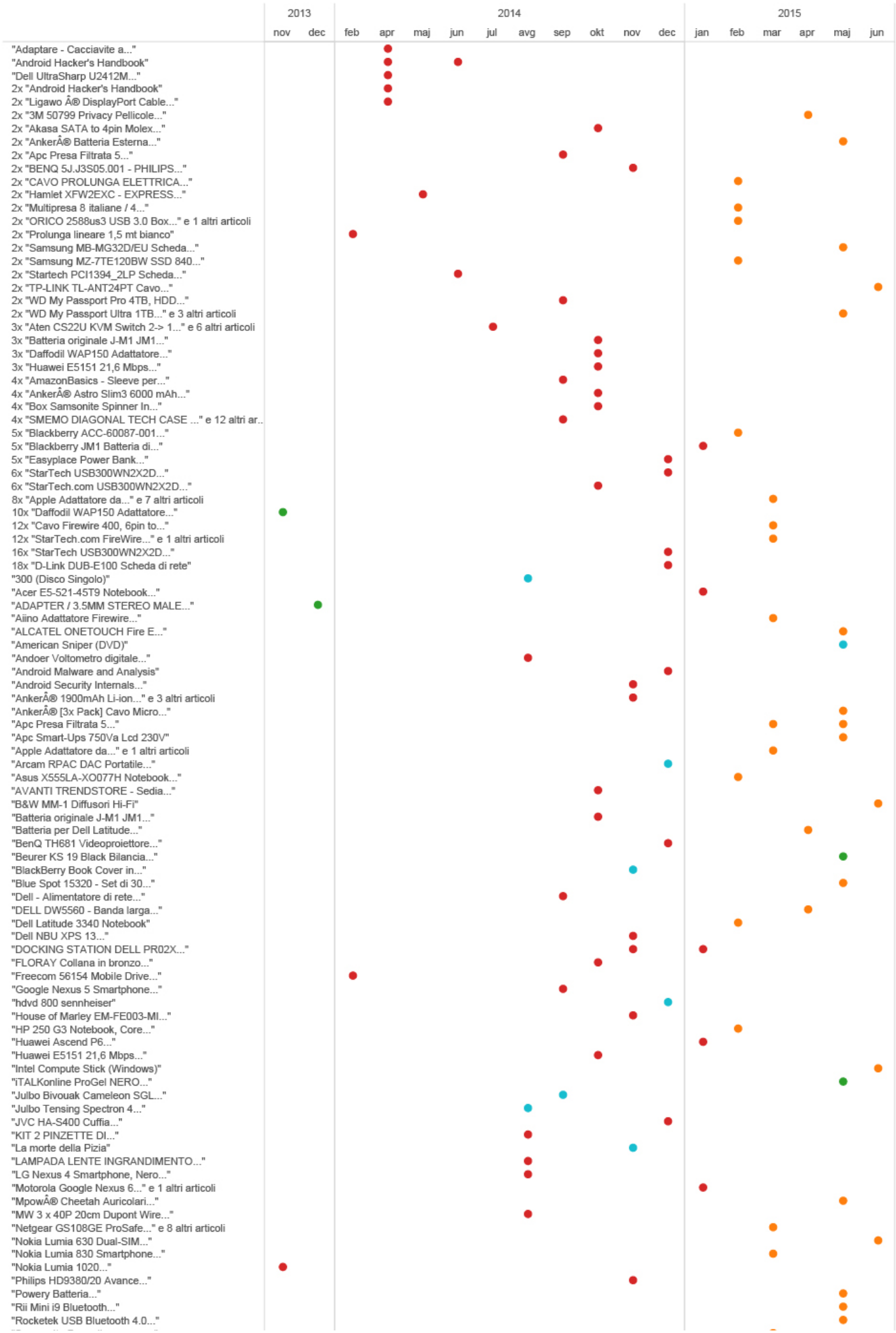
TOOL : TABLEAU

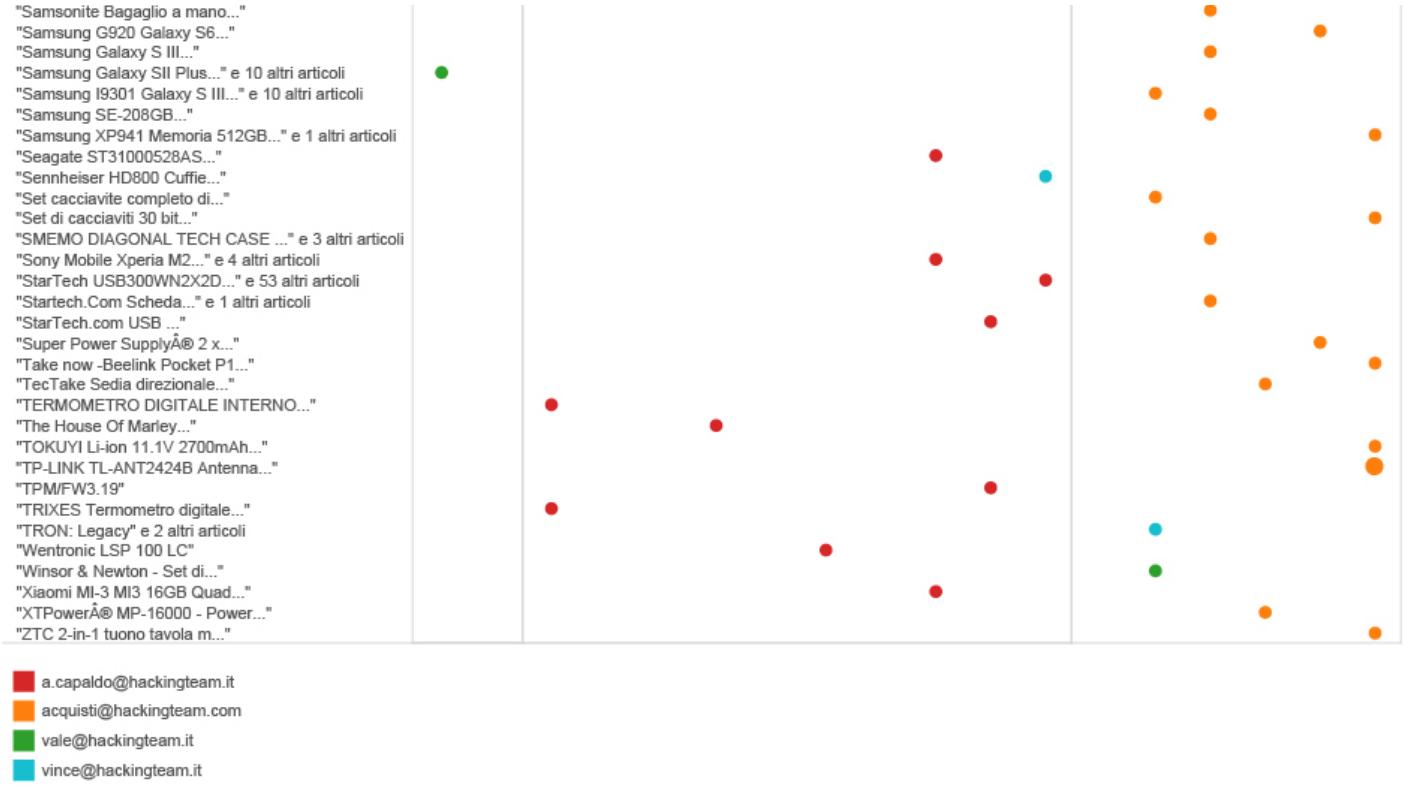
Getting back to email subjects, they can sometimes give us some really amusing information. For instance, many companies such as Amazon, list the ordered items in the subject of a Confirmation email you receive after

your payment has been processed.

By extracting the emails sent by Amazon to Hacking Team employees, we were able to get an insight into their purchases.

TIMELINE OF EMAILS WITH SUBJECTS FROM AMAZON.IT


2013
nov
dec
2014
feb
apr
maj
jun
jul
avg
sep
okt
nov
dec
2015
jan
feb
mar
apr
maj
jun
"Adaptare - Cacciavite a..."
"Android Hacker's Handbook"
"Dell UltraSharp U2412M..."
2x "Android Hacker's Handbook"
2x "Ligawo Â® DisplayPort Cable..."
2x "3M 50799 Privacy Pellicole..."
2x "Akasa SATA to 4pin Molex..."
2x "AnkerÂ® Batteria Esterna..."
2x "Apc Presa Filtrata 5..."
2x "BENQ 5J.J3S05.001 - PHILIPS..."
2x "CAVO PROLUNGA ELETTRICA..."
2x "Hamlet XFW2EXC - EXPRESS..."
2x "Multipresa 8 italiane / 4..."
2x "ORICO 2588us3 USB 3.0 Box..." e 1 altri articoli
2x "Prolunga lineare 1,5 mt bianco"
2x "Samsung MB-MG32D/EU Scheda..."
2x "Samsung MZ-7TE120BW SSD 840..."
2x "Startech PCI1394_2LP Scheda..."
2x "TP-LINK TL-ANT24PT Cavo..."
2x "WD My Passport Pro 4TB, HDD..."
2x "WD My Passport Ultra 1TB..." e 3 altri articoli
3x "Aten CS22U KVM Switch 2-> 1..." e 6 altri articoli
3x "Batteria originale J-M1 JM1..."
3x "Daffodil WAP150 Adattatore..."
3x "Huawei E5151 21,6 Mbps..."
4x "AmazonBasics - Sleeve per..."
4x "AnkerÂ® Astro Slim3 6000 mAh..."
4x "Box Samsonite Spinner In..."
4x "SMEMO DIAGONAL TECH CASE ..." e 12 altri ar..
5x "Blackberry ACC-60087-001..."
5x "Blackberry JM1 Batteria di..."
5x "Easyplace Power Bank..."
6x "StarTech USB300WN2X2D..."
6x "StarTech.com USB300WN2X2D..."
8x "Apple Adattatore da..." e 7 altri articoli
10x "Daffodil WAP150 Adattatore..."
12x "Cavo Firewire 400, 6pin to..."
12x "StarTech.com FireWire..." e 1 altri articoli
16x "StarTech USB300WN2X2D..."
18x "D-Link DUB-E100 Scheda di rete"
"300 (Disco Singolo)"
"Acer E5-521-45T9 Notebook..."
"ADAPTER / 3.5MM STEREO MALE..."
"Aiino Adattatore Firewire..."
"ALCATEL ONETOUCH Fire E..."
"American Sniper (DVD)"
"Andoer Voltometro digitale..."
"Android Malware and Analysis"
"Android Security Internals..."
"AnkerÂ® 1900mAh Li-ion..." e 3 altri articoli
"AnkerÂ® [3x Pack] Cavo Micro..."
"Apc Presa Filtrata 5..."
"Apc Smart-Ups 750Va Lcd 230V"
"Apple Adattatore da..." e 1 altri articoli
"Arcam RPAC DAC Portatile..."
"Asus X555LA-XO077H Notebook..."
"AVANTI TRENDSTORE - Sedia..."
"B&W MM-1 Diffusori Hi-Fi"
"Batteria originale J-M1 JM1..."
"Batteria per Dell Latitude..."
"BenQ TH681 Videoproiettore..."
"Beurer KS 19 Black Bilancia..."
"BlackBerry Book Cover in..."
"Blue Spot 15320 - Set di 30..."
"Dell - Alimentatore di rete..."
"DELL DW5560 - Banda larga..."
"Dell Latitude 3340 Notebook"
"Dell NBU XPS 13..."
"DOCKING STATION DELL PR02X..."
"FLORAY Collana in bronzo..."
"Freecom 56154 Mobile Drive..."
"Google Nexus 5 Smartphone..."
"hdvd 800 sennheiser"
"House of Marley EM-FE003-MI..."
"HP 250 G3 Notebook, Core..."
"Huawei Ascend P6..."
"Huawei E5151 21,6 Mbps..."
"Intel Compute Stick (Windows)"
"iTALKonline ProGel NERO..."
"Julbo Bivouak Cameleon SGL..."
"Julbo Tensing Spectron 4..."
"JVC HA-S400 Cuffia..."
"KIT 2 PINZETTE DI..."
"La morte della Pizia"
"LAMPADA LENTE INGRANDIMENTO..."
"LG Nexus 4 Smartphone, Nero..."
"Motorola Google Nexus 6..." e 1 altri articoli
"MpowÂ® Cheetah Auricolari..."
"MW 3 x 40P 20cm Dupont Wire..."
"Netgear GS108GE ProSafe..." e 8 altri articoli
"Nokia Lumia 630 Dual-SIM..."
"Nokia Lumia 830 Smartphone..."
"Nokia Lumia 1020..."
"Philips HD9380/20 Avance..."
"Powery Batteria..."
"Rii Mini i9 Bluetooth..."
"Rocketek USB Bluetooth 4.0..."



But there are some more extreme examples.

If we look at the treemap of Hacking Team partners, there is a company called Carlson Wagonit Travel. According to the company's website, they deliver solutions for business travel, meetings and events management. For HT they arrange and buy plane tickets, book hotels and provide travel assistance. They have one bad habit (which is quite common for many booking agencies), every time an airplane ticket is booked, the agency would send an email with name and airport codes, contained in the subject line, to the prospective passenger. Extracting that information from the subject and cross-referencing with the date the email was sent, we are able to get an approximate information about the journeys of HT employees.

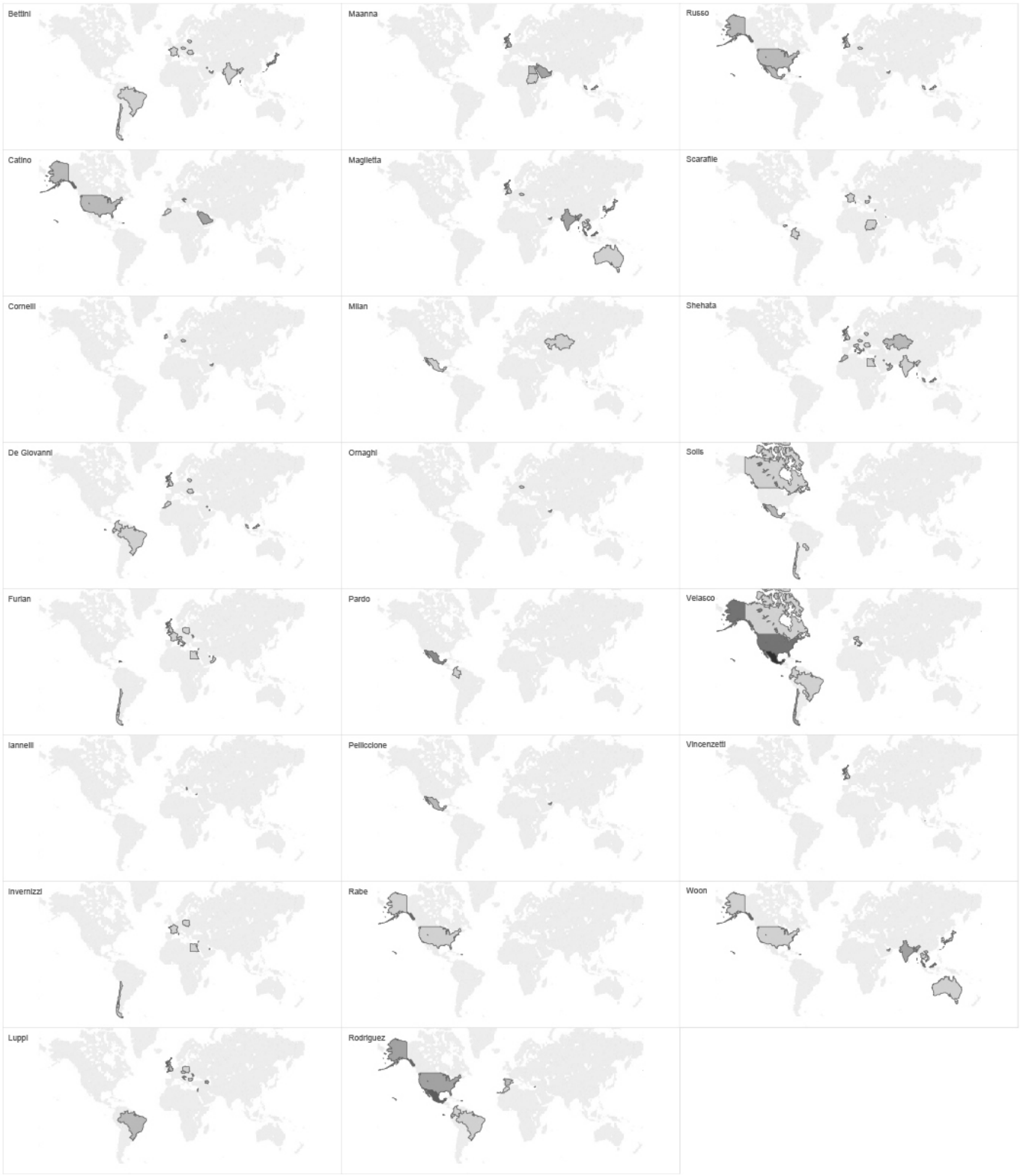
The list of Hacking Team frequent flyers and locations they visit looks like this.

MAP OF HT EMPLOYEES FLIGHTS BASED ON CWT EMAILS SUBJECT LINES

TOOL : TABLEAU

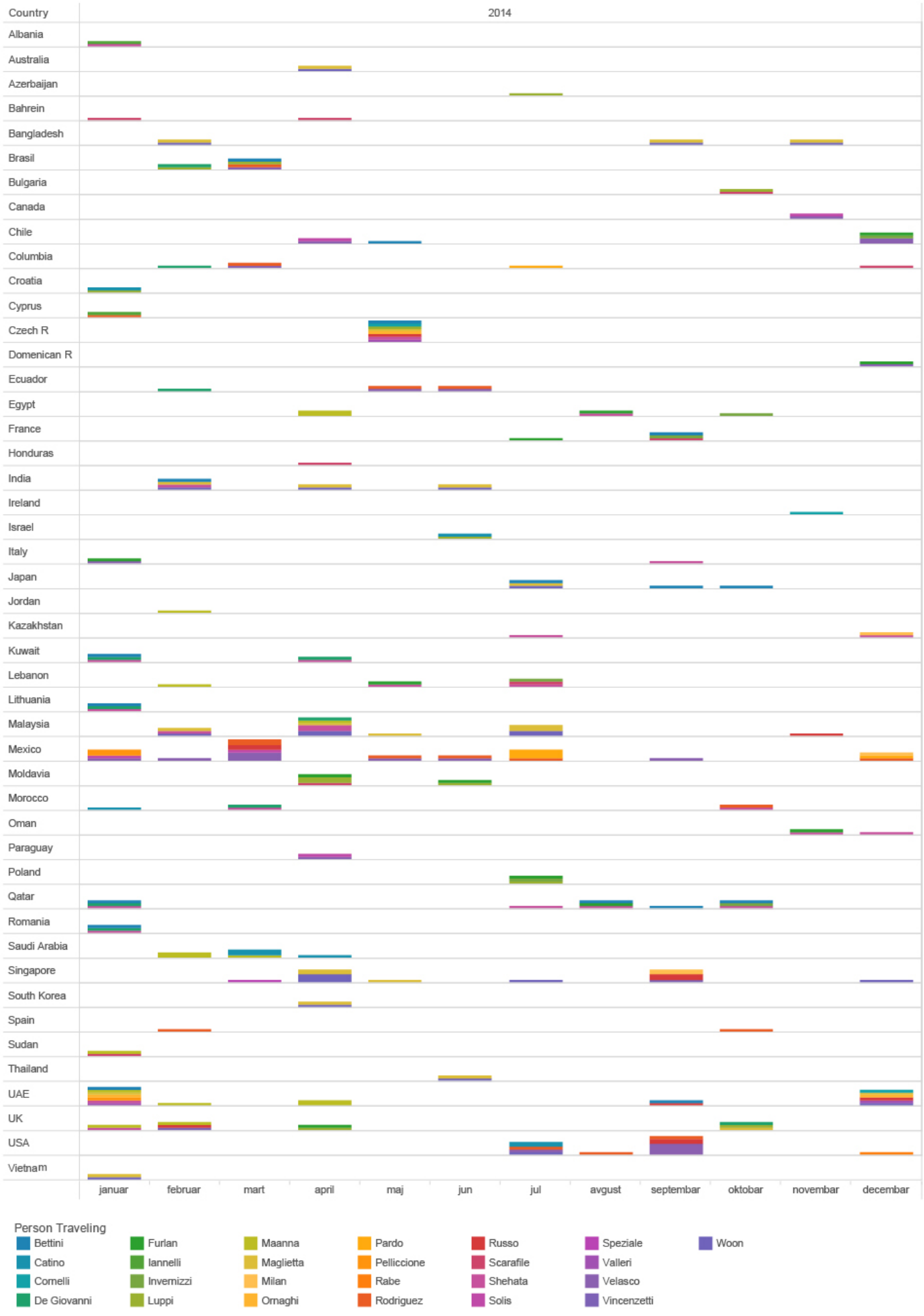
These data give some interesting information related to our assumption of how the organisational structure of Hacking Team looks like. If we go back to our organisational structure graph based on social network analysis and compare with this list of frequent flyers, we can see that the most frequent flyers are not very high in the hierarchy according to our network graph. However, if we group the flights by passenger's name, we realise that each of the most frequent flyers is based in a certain place, and covers a certain region/market, such as SE Asia, Middle East, South America etc. Conclusion that we can get from this is that those employees are responsible for certain markets or regional Hacking team offices around the world.

INDIVIDUAL HT EMPLOYEES FLIGHTS MAP BASED ON CWT EMAILS SUBJECT LINES



If we regroup the same set of data, by location, we can see at which point in time and where two or more Hacking Team employees have met or have traveled together. This implies potential business meetings, sales of surveillance tools, establishing new relations with international customers and government agencies around the globe.

TIMELINE OF INDIVIDUAL HT EMPLOYEES FLIGHTS TO DIFFERENT COUNTRIES BASED ON CWT EMAILS SUBJECT LINES


Country
2014
Albania
Australia
Azerbaijan
Bahrein
Bangladesh
Brasil
Bulgaria
Canada
Chile
Columbia
Croatia
Cyprus
Czech R
Domenican R
Ecuador
Egypt
France
Honduras
India
Ireland
Israel
Italy
Japan
Jordan
Kazakhstan
Kuwait
Lebanon
Lithuania
Malaysia
Mexico
Moldavia
Morocco
Oman
Paraguay
Poland
Qatar
Romania
Saudi Arabia
Singapore
South Korea
Spain
Sudan
Thailand
UAE
UK
USA
Vietnam
januar
februar
mart
april
maj
jun
jul
avgust
septembar
oktobar
novembar
decembar
Person Traveling
Bettini
Catino
Cornelli
De Giovanni
Furlan
Iannelli
Invernizzi
Luppi
Maanna
Maglietta
Milan
Ornaghi
Pardo
Pelliccione
Rabe
Rodriguez
Russo
Scarafile
Shehata
Solis
Speziale
Valleri
Velasco
Vincenzetti
Woon

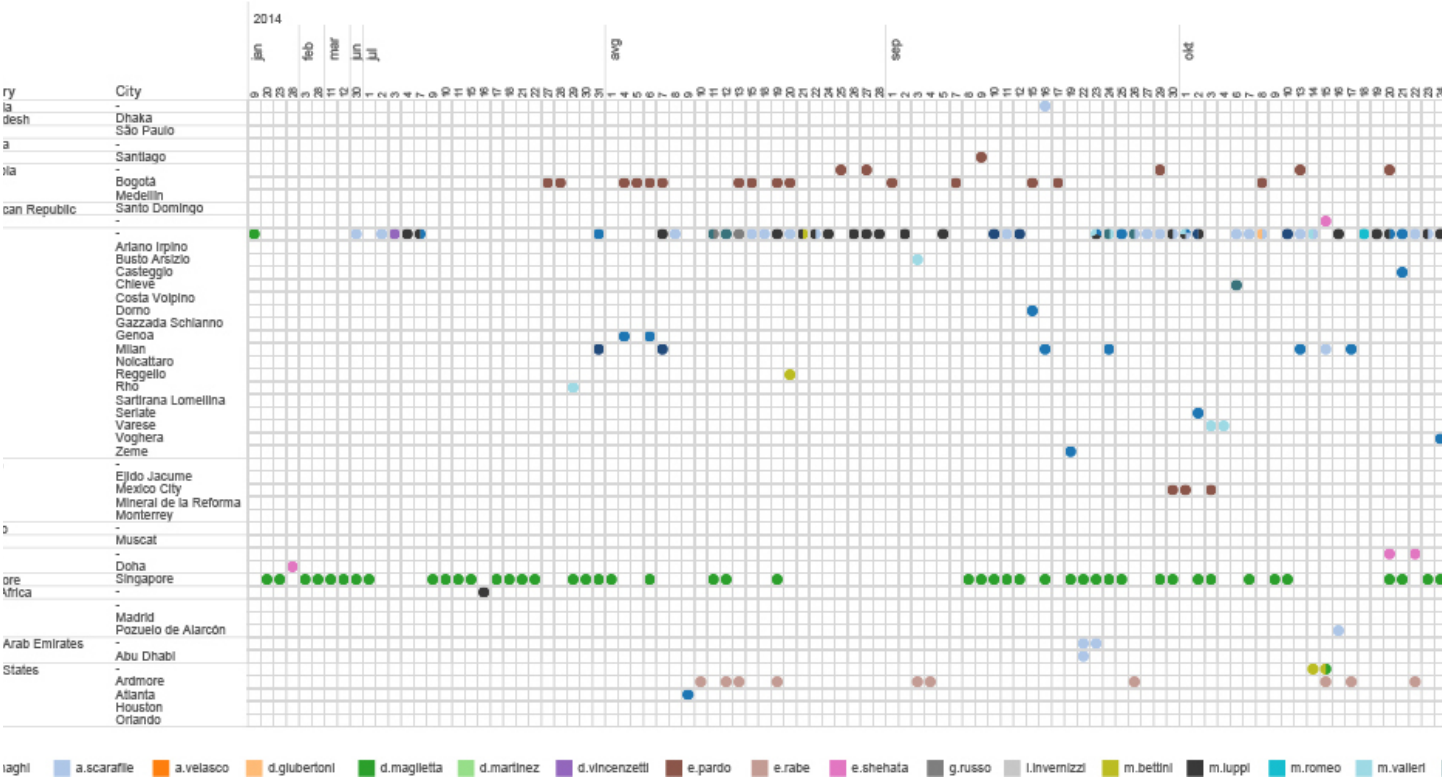
TOOL : TABLEAU

HOMING PIGEON

The email header hides one even more precise location information. In some cases, the email headers reveal the IP address of the sender. The IP address can then be geolocated, using some publicly available tools, to the level of a city or individual router. In the context of our investigation, this allows us to trace every one of Mr.D's contacts. Every time someone sends an email to Mr.D, that person basically reveals their location to us. Just by analysing the metadata of Mr.D's received emails we can get information where the senders are located, when they changed city or country.

This allowed us to locate even more precisely in time all employees of Hacking Team.

TIMELINE OF HT EMPLOYEES IP LOCATIONS BASED ON EMAILS RECEIVED BY MR.D (2014-2015)



We can see for example, that most of the employees are located in Italy, but there are some of them being situated in different places around the world.

D.Maglietta is for example head of their office in Singapore but he spends Christmas in Spain. **E.Pardo** is in Bogota, Colombia most of the time and he had a visit from another HT employee **A.Scarafile** on January 14th. We can see how **E.Shehata** jumps around the Middle East visiting Qatar, Lebanon, United Arab Emirates, Egypt and Jordan. **E. Rabe** is situated in Ardmore, PA, USA but moved on February 14th to another suburb of Philadelphia. And finally, even though HT claims to have an office in Washington DC, based on metadata we didn't find any evidence that would support that.

On a World map, the distribution of their locations looks like this.

MAP OF HT EMPLOYEES IP LOCATIONS BASED ON EMAILS RECEIVED BY MR.D (2014-2015)

TOOL : TABLEAU

Locations of the external contacts of Mr.D give us a real insight into their global operations.

MAP OF EXTERNAL CONTACTS IP LOCATIONS BASED ON EMAILS RECEIVED BY MR.D (2014-2015)

TOOL : TABLEAU

This tiny piece of information allows us to explore individual contacts in each country, to find their main partners, even to track locations of their contacts in time.

MAP OF EXTERNAL CONTACTS IP LOCATIONS (ZOOM OR MOVE MAP TO EXPLORE OTHER REGIONS)

TOOL : TABLEAU

IT'S JUST METADATA

More often than not, the power of metadata is being taken naively or its potential usage is being oversimplified in comparison with the content of our communication. But we see that even our not very sophisticated, DIY methods, enabled us to create a deep and clear image of someone's habits and activities, using information extracted from 'only' email metadata. Although our investigation primarily discovered relations, patterns and anomalies of someone's work life, it still gave us an insight into that person's habits that border with private life. In the end, metadata scans someone's behaviors on a much deeper level than traditional surveillance practice related to content could ever do.

At moments, while conducting this investigation, it certainly felt as if we were peeking into the deepest corners of someone's life. What felt even more disturbing is the idea that our subjects of analysis are probably less aware of their behavioral patterns that we are. It's just metadata, and in our case just one little segment of it. This is why our research provoked an internal debate within our team on the ethical issues of this kind of practice and on the form in which the findings of this research should be published. On numerous occasions, supporters of NSA surveillance programs, claimed that collecting and analysis of metadata is not surveillance. According to our data exploration, we can claim that it can be even more intrusive than regular content surveillance.

WHO HAS ACCESS TO METADATA?

Understanding who has access to metadata and the possibility to analyse it will give us an answer to the question of the new power structures and

distribution of wealth in the information society.

COMPANIES

THE FIRST AND OBVIOUS GROUP ARE THE COMPANIES THAT PROVIDE SERVICES SUCH AS GOOGLE OR FACEBOOK. THEY DON'T JUST HAVE ACCESS TO THE METADATA, THEY HAVE ACTUAL DATA AND CONTENT ON THEIR SERVERS. QUALITY, VARIETY AND AMOUNT OF METADATA THAT FOR EXAMPLE GOOGLE OWNS ABOUT EVERY USER OF HIS SERVICES, STORED IN GIGANTIC DATACENTRES ARE UNPRECEDENTED MONOPOLIES OF COLLECTIVE DATA. THROUGH THEIR CORE BUSINESS MODELS THEY ARE PIONEERS OF METADATA EXPLOITATION. IT COULD EVEN BE SAID THAT THE FIRST DATA CENTRE SETUP BY GOOGLE IN 1998 CAN BE CONSIDERED THE MILESTONE OF THE BIRTH OF THE METADATA SOCIETY.

INFRASTRUCTURE

THE SECOND GROUP WITH ACCESS TO METADATA IS RELATED TO THE THE INTERNET'S INFRASTRUCTURE. THOSE ARE INTERNET SERVICE PROVIDERS, MOBILE SERVICE PROVIDERS, INTERNET EXCHANGE POINTS AND SUBMARINE OPTIC CABLES AND THEY CAN ACCESS DATA WHEN IT FLOWS THROUGH THEIR CABLES, ROUTERS AND SERVERS. THE QUALITY OF METADATA THAT THEY CAN COLLECT DEPENDS ON THE ENDPOINTS ATTACHED TO THE INFRASTRUCTURE, BUT IN ANY CASE THEY HAVE ACCESS TO THE BASIC METADATA OF THE INTERNET PACKETS. EVEN THOUGH THE CONTENT OF, SAY, EMAILS, IS SUPPOSEDLY ENCRYPTED (AT LEAST WHEN USING THESE MAINSTREAM SERVICES), THE METADATA IS NOT, BECAUSE THE EMAIL ARCHITECTURE SIMPLY HAS TO RELY ON METADATA THAT ARE READABLE, AS EXPLAINED BEFORE.

GOVERNMENT

NATIONAL LAWS IN MOST CASES GIVE THE GOVERNMENT OR SOME OF THE AGENCIES THEREOF A LEGITIMATE ACCESS TO USERS' DATA, INCLUDING EMAIL AND OTHER METADATA. INTERNET INFRASTRUCTURE OWNERS OR COMPANIES THAT PROVIDE SERVICES ARE OBLIGED TO COOPERATE WITH GOVERNMENTS WHOSE JURISDICTION THEY ARE UNDER. THEY OFTEN COMPLY WITH GOVERNMENT REQUESTS AND HAVE DIFFERENT FORMS OF TECHNICAL COOPERATIONS. HOWEVER, IN MANY COUNTRIES GOVERNMENT AGENCIES HAVE INVESTED IN DEVELOPING PROGRAMS FOR MASS SURVEILLANCE OF CITIZENS BASED ON COLLECTING METADATA. THESE PROGRAMS OFTEN RELY ON SOFTWARE THAT CREATES AND EXPLOITS BACKDOORS (AS IS HACKING TEAM'S CASE) OR USE SOME OTHER CREATIVE WAY TO GET ACCESS TO METADATA.

Those are the hunters, hoarders and scavengers in the ecosystem of the metadata society, but there are some smaller species worth mentioning as well. Metadata is often a resource for different businesses based on data collection and analysis in the field of digital marketing, business analytics or scientific research. You can even be a subject of surveillance in your work environment. The company you work for could perform metadata

analysis of your productivity, anomalies in your behavior during work time and analysis of your contacts. To name an example, this kind of internal company surveillance service is provided by one of the Hacking Team main partners – Nice, mentioned earlier in this text.

Understanding the power of metadata brings us closer to understanding the algorithmic governmentality as a concept and practice. The quality of metadata is that it is really easy to process and it can be done by machines and algorithms. In the eye of the algorithm, we are observed through our profiles, sets of behavioral patterns and anomalies extracted from our metadata. Automatic processing and algorithmic analysis of those data in real time leads to the world in which algorithms can decide whether we are terrorists or regular citizens, are we suitable for a loan in the bank, an insurance policy, or who is going to appear in our social stream. Algorithms can eventually predict our future behaviour based on our past metadata, bringing us closer to the concept of pre-crime, the tendency in criminal justice systems to focus on crimes not yet committed.

WE ARE NOT GOING TO CONCLUDE ANYTHING ABOUT HACKING TEAM'S ACTIVITIES, BECAUSE THAT WAS NEVER THE GOAL OF OUR RESEARCH. WE WANTED TO UNDERSTAND, HANDS-ON, THROUGH RESEARCH AND PRACTICE HOW METADATA ANALYSIS CAN BE PERFORMED AND WHAT WE CAN LEARN FROM IT. WE HOPE THAT RESEARCHERS AND INVESTIGATIVE JOURNALISTS CAN USE OUR DATA AND EXPLORATION FOR THEIR OWN RESEARCH AND THAT THEY WILL BE ABLE TO FIND

NEW CONNECTIONS AND LEADS BASED ON METADATA.

Contributors

VLADAN JOLER – CONCEPT, RESEARCH, DATA ANALYSIS, DATA VISUALIZATION, TEXT

ANDREJ PETROVSKI – DATA MINING AND PROCESSING, TEXT

NIKOLA KOTUR – DATA PROCESSING AND DEVELOPMENT

TAMARA PAVLOVIC – EDITING AND PROOFREADING

JAN KRASNI – PEER REVIEW

MORE ABOUT SHARE LAB YOU CAN FIND HERE

FOR ANY QUESTION, RAW DATA FOR RESEARCH, OR ANYTHING ELSE, PLEASE CONTACT US ON INFO.AT.SHARECONFERENCE.NET

PREVIOUS POST

Category: Investigating metadata

SHARE ON

TWITTER FACEBOOK

About
Raw Data, Documents & Tools
Terms and Conditions & Privacy Policy
Contact us
English
Srpski

RECENT POSTS

Metadata Investigation : Inside Hacking Team

Hacking Team : The “Italian job” of Serbian security services

Invisible Infrastructures : Surveillance Achitecture

Invisible Infrastructures : Data Flow

Invisible Infrastructures : Online Trackers

About / Raw Data, Documents & Tools / Terms and Conditions & Privacy Policy / Contact us / English / Srpski