Tracing Cross Border Web Tracking

Costas Iordanou Georgios Smaragdakis
Ingmar Poese Nikolaos Laoutaris

Technische Universität Berlin
Universidad Carlos III de Madrid
BENOCS
DATA TRANSPARENCY LAB
erc European Research Council

General Data Protection Regulation - Details

One of the biggest changes with respect to privacy and regulation on the web in the last few years (Enforcement date: 25th May, 2018)

In general the new regulation:

1. tries to regulate how users' data are collected, processed and stored and
2. if they include any sensitive information about the user

General Data Protection Regulation - Details

One of the biggest changes with respect to privacy and regulation on the web in the last few years (Enforcement date: 25th May, 2018)

In general the new regulation:

1. tries to regulate how users' data are collected, processed and stored and
2. if they include any sensitive information about the user

Implementation – Per member state Data Protection Authority (DPA)

DPA: Responsible for complains – investigations & enforcement

Investigation starting point – **Ad & Tracking** flows entry point servers location

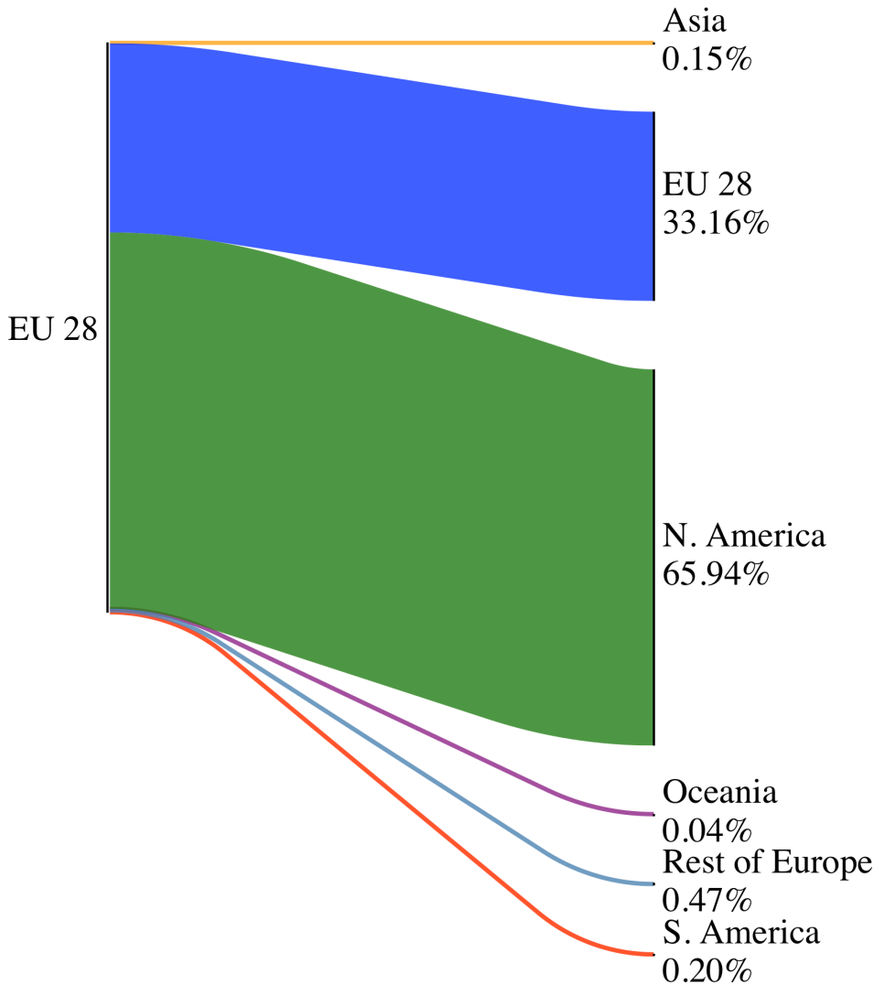
RQ: How can we identify the physical locations of such servers?

How to monitor backend tracking servers location?

By following a four step methodology and avoiding pitfalls:

1. Use **real users** from **different geographic locations**
2. Map 3rd party **domains to IPs**
3. **Identify Ad and Tracking** related domains.
4. Correctly **geo-locate** the backend server

What is the EU 28 confinement level?



MaxMind geo-location

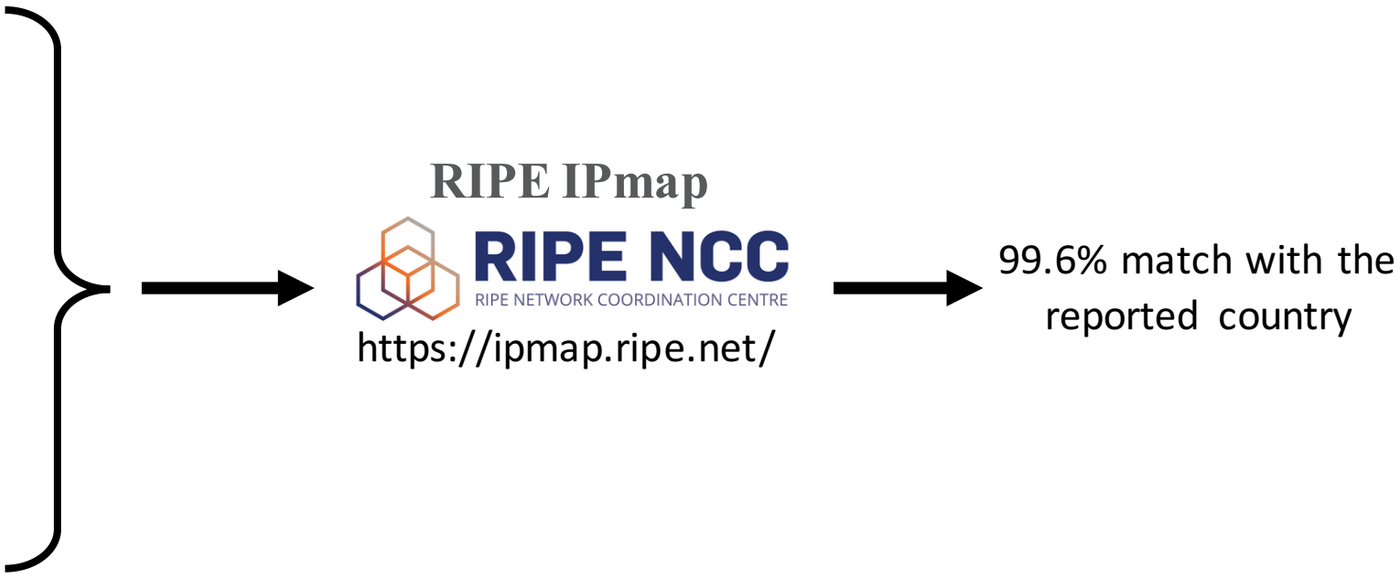
Accurate geo-location of server IPs

RIPE IPmap validation process - infrastructure servers IPs

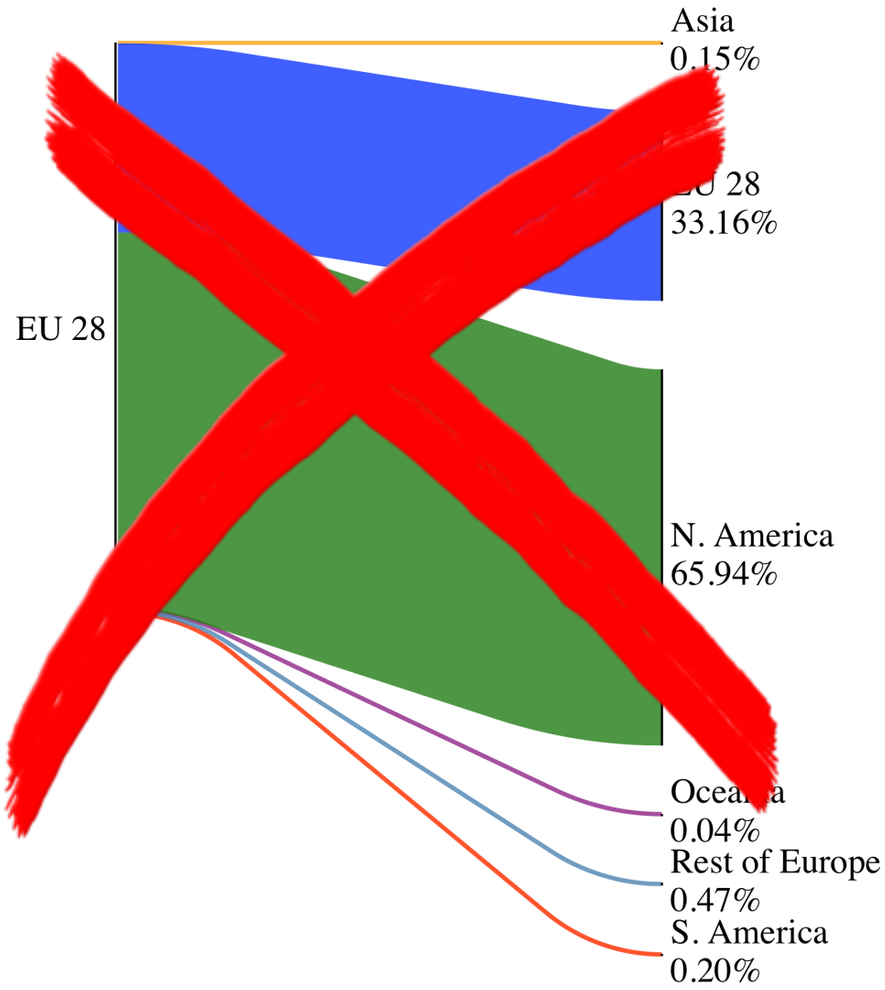
prefix	region	service
46.51.128.0/18	eu-west-1	AMAZON
46.51.216.0/21	ap-southeast-1	AMAZON
13.73.232.0/21	japaneast	AZURE
20.190.144.128/25	koreacentral	AZURE
...	...	...

Regions map

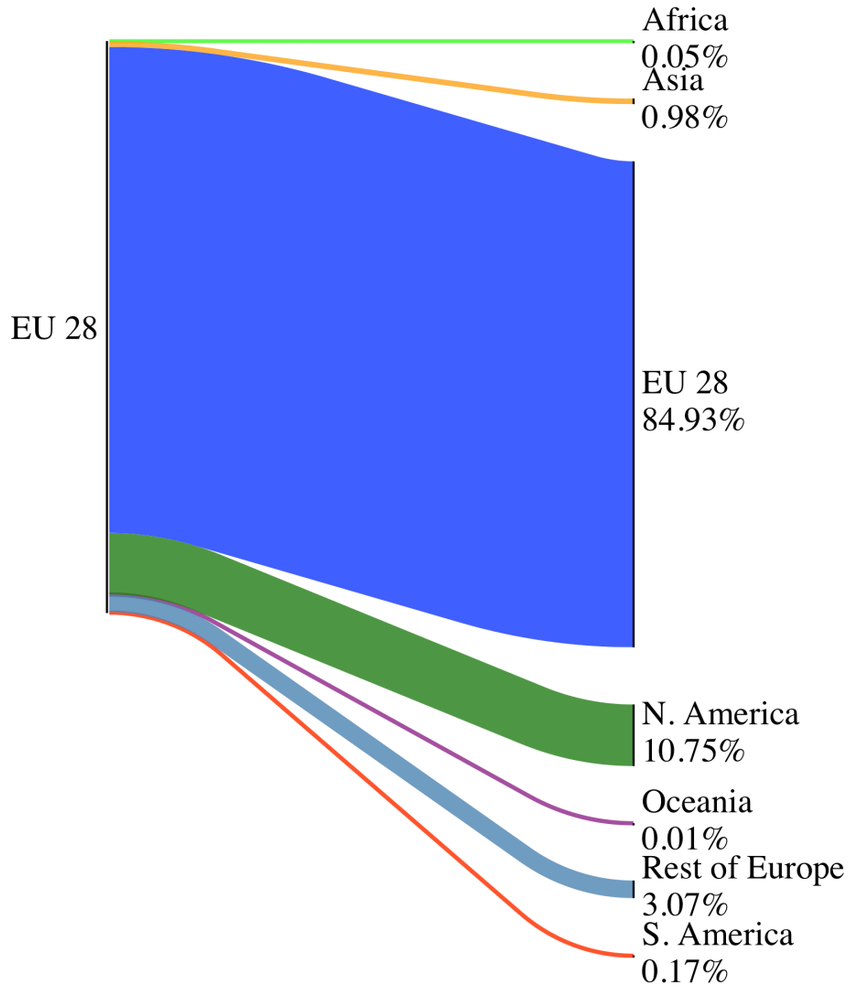
eu-west-1: Ireland, Ireland
ap-southeast-1: Singapore, Singapore



What is the EU 28 confinement level?



MaxMind geo-location



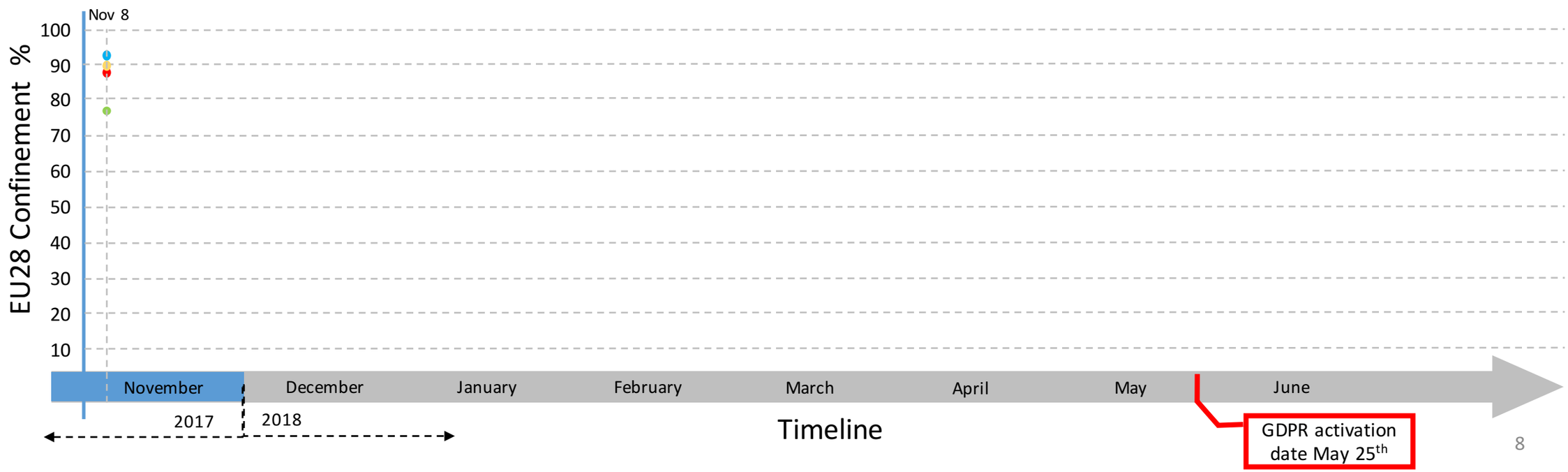
RIPE IPmap geo-location

Scaling up – Continent level ISPs results

	● DE-Broadband				● DE-Mobile				● PL				● HU			
	Nov 8				Nov 8				Nov 8				Nov 8			
#Sampled Tracking Flows (in Millions)	1,057.0				70.4				13.8				43.3			
EU28	**88.5%**				**91.1%**				**77.5%**				**89.5%**			
North America	10%				6.9%				19.8%				10.2%			
Rest Europe	<1%				<1%				1.9%				<1%			
Asia	<1%				<1%				<1%				<1%			
Rest World	<1%				<1%				<1%				<1%			

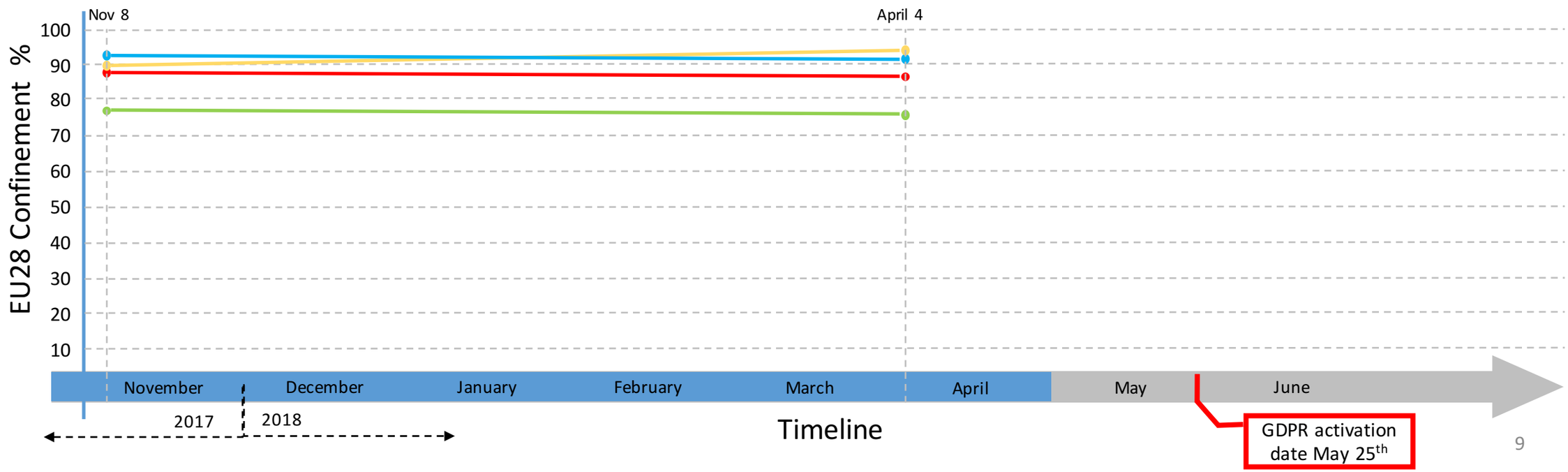
Scaling up – Continent level ISPs results

	● DE-Broadband				● DE-Mobile				● PL				● HU			
	Nov 8				Nov 8				Nov 8				Nov 8			
#Sampled Tracking Flows (in Millions)	1,057.0				70.4				13.8				43.3			
EU28	**88.5%**				**91.1%**				**77.5%**				**89.5%**			
North America	10%				6.9%				19.8%				10.2%			
Rest Europe	<1%				<1%				1.9%				<1%			
Asia	<1%				<1%				<1%				<1%			
Rest World	<1%				<1%				<1%				<1%			



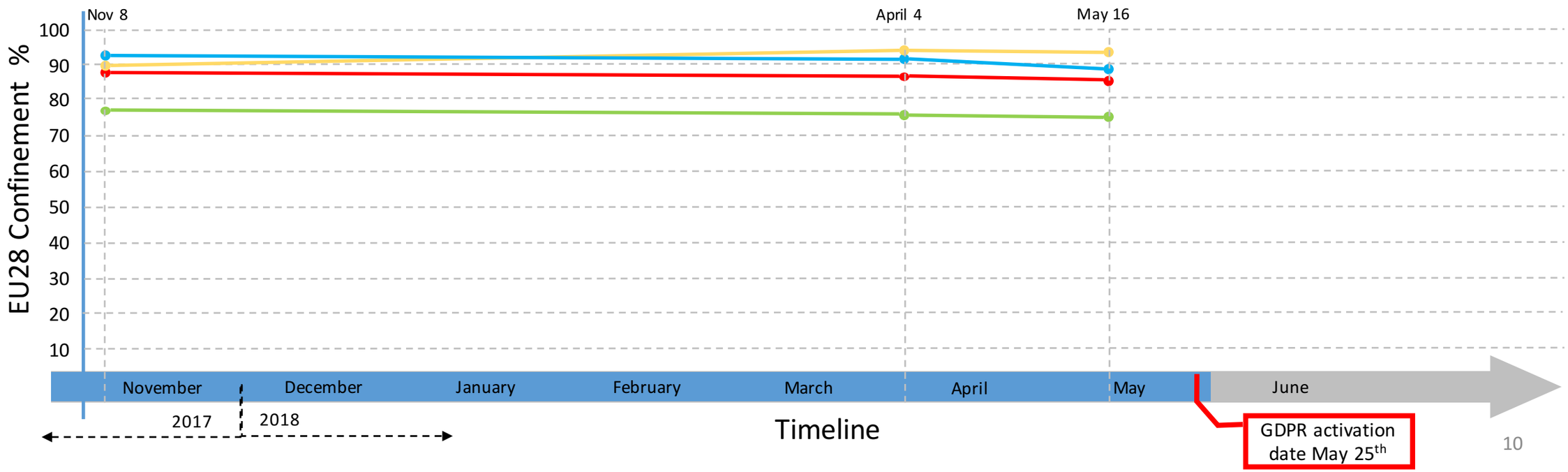
Scaling up – Continent level ISPs results

	DE-Broadband				DE-Mobile				PL				HU			
	Nov 8	April 4			Nov 8	April 4			Nov 8	April 4			Nov 8	April 4		
#Sampled Tracking Flows (in Millions)	1,057.0	1,200.8			70.4	77.4			13.8	13.8			43.3	50.2		
EU28	**88.5%**	**87.7%**			**91.1%**	**90.8%**			**77.5%**	**75.6%**			**89.5%**	**93.1%**		
North America	10%	9.3%			6.9%	6.6%			19.8%	21.5%			10.2%	6.3%		
Rest Europe	<1%	1.7%			<1%	2%			1.9%	1.9%			<1%	<1%		
Asia	<1%	<1%			<1%	<1%			<1%	<1%			<1%	<1%		
Rest World	<1%	<1%			<1%	<1%			<1%	<1%			<1%	<1%		



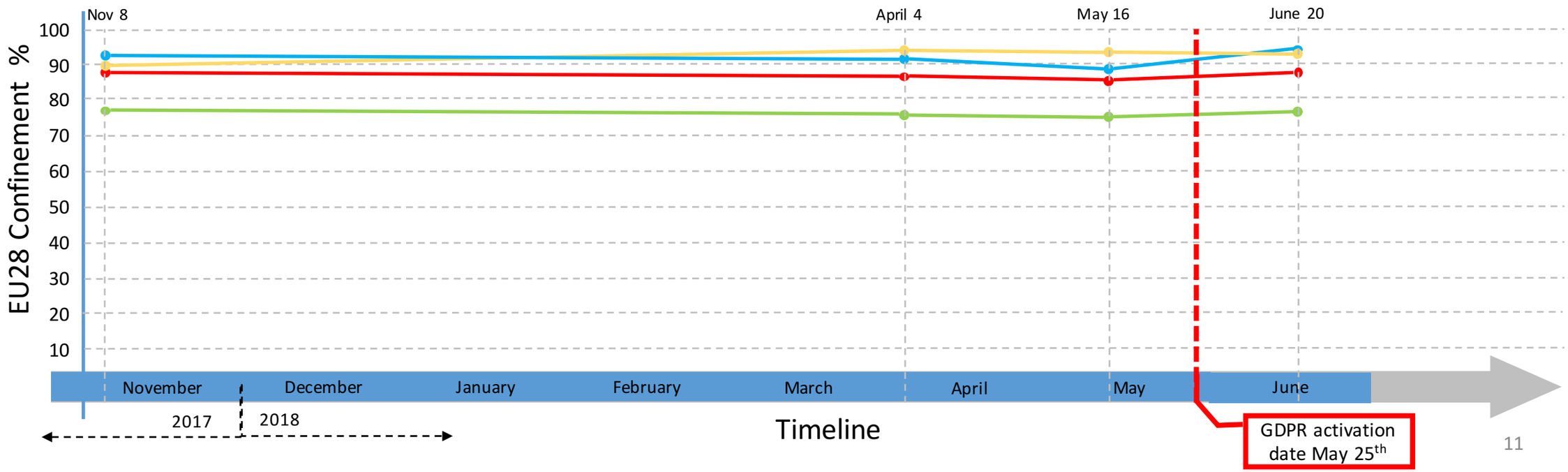
Scaling up – Continent level ISPs results

	DE-Broadband				DE-Mobile				PL				HU			
	Nov 8	April 4	May 16		Nov 8	April 4	May 16		Nov 8	April 4	May 16		Nov 8	April 4	May 16	
#Sampled Tracking Flows (in Millions)	1,057.0	1,200.8	1,105.3		70.4	77.4	70.8		13.8	13.8	12.4		43.3	50.2	39.3	
EU28	**88.5%**	**87.7%**	**86.5%**		**91.1%**	**90.8%**	**89.9%**		**77.5%**	**75.6%**	**74.7%**		**89.5%**	**93.1%**	**92.4%**	
North America	10%	9.3%	9.2%		6.9%	6.6%	6.4%		19.8%	21.5%	22%		10.2%	6.3%	7%	
Rest Europe	<1%	1.7%	2.9%		<1%	2%	3.1%		1.9%	1.9%	1.7%		<1%	<1%	<1%	
Asia	<1%	<1%	<1%		<1%	<1%	<1%		<1%	<1%	<1%		<1%	<1%	<1%	
Rest World	<1%	<1%	<1%		<1%	<1%	<1%		<1%	<1%	1.1%		<1%	<1%	<1%	



Scaling up – Continent level ISPs results

	DE-Broadband				DE-Mobile				PL				HU			
	Nov 8	April 4	May 16	June 20	Nov 8	April 4	May 16	June 20	Nov 8	April 4	May 16	June 20	Nov 8	April 4	May 16	June 20
#Sampled Tracking Flows (in Millions)	1,057.0	1,200.8	1,105.3	963.4	70.4	77.4	70.8	74.5	13.8	13.8	12.4	11.9	43.3	50.2	39.3	33.6
EU28	**88.5%**	**87.7%**	**86.5%**	**88.3%**	**91.1%**	**90.8%**	**89.9%**	**92.5%**	**77.5%**	**75.6%**	**74.7%**	**75%**	**89.5%**	**93.1%**	**92.4%**	**91.6%**
North America	10%	9.3%	9.2%	8.4%	6.9%	6.6%	6.4%	5.1%	19.8%	21.5%	22%	21.3%	10.2%	6.3%	7%	7.7%
Rest Europe	<1%	1.7%	2.9%	1.8%	<1%	2%	3.1%	1.3%	1.9%	1.9%	1.7%	3.4%	<1%	<1%	<1%	<1%
Asia	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%
Rest World	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	1.1%	<1%	<1%	<1%	<1%	<1%



In the paper (to appear at IMC 2018)

1. More details on the methodology
2. Confinement improvement suggestions
3. More results

Tracing Cross Border Web Tracking

Costas Iordanou
TU Berlin / UC3M
costas@ima.tu-berlin.de

Georgios Smaragdakis
TU Berlin
georgios@ima.tu-berlin.de

Ingmar Poese
BENOCS
ipoese@benocs.com

Nikolaos Laoutaris
Data Transparency Lab / Eurecat
nikos@datatransparencylab.org

ABSTRACT

A tracking flow is a flow between an end user and a Web tracking service. We develop an extensive measurement methodology for quantifying at scale the amount of tracking flows that cross data protection borders, be it national or international, such as the EU28 border within which the General Data Protection Regulation (GDPR) applies. Our methodology uses a browser extension to fully render advertising and tracking code, various lists and heuristics to extract well known trackers, passive DNS replication to get all the IP ranges of trackers, and state-of-the art geolocation. We employ

1 INTRODUCTION

Online advertising, including bahavioral targeting over the Real Time Bidding protocol (RTB) [62], fuels [26] most of the free services of the web. In its principle, the concept of targeted (or personalized) advertising is benign: products and services offered to consumers that they truly care about. It is in its implementation and actual use when controversies arise. For example, tracking should respect fundamental data protection rights of people, such as their desire to opt-out, and should keep clear from sensitive personal data categories, such as health, political beliefs, religion or sexual

Tracing Cross Border Web Tracking

Costas Iordanou
email: costas@ima.tu-berlin.de

Technische Universität Berlin

Universidad Carlos III de Madrid

BENOCS

DATA TRANSPARENCY LAB

erc European Research Council