Semantic Exploration of DNS

Samuel Marchal, Jérôme François, Cynthia Wagner and Thomas Engel

UNIVERSITÉ DU LUXEMBOURG

SnT

NETLAB

SNT

Outline

SMT

Outline

SNT

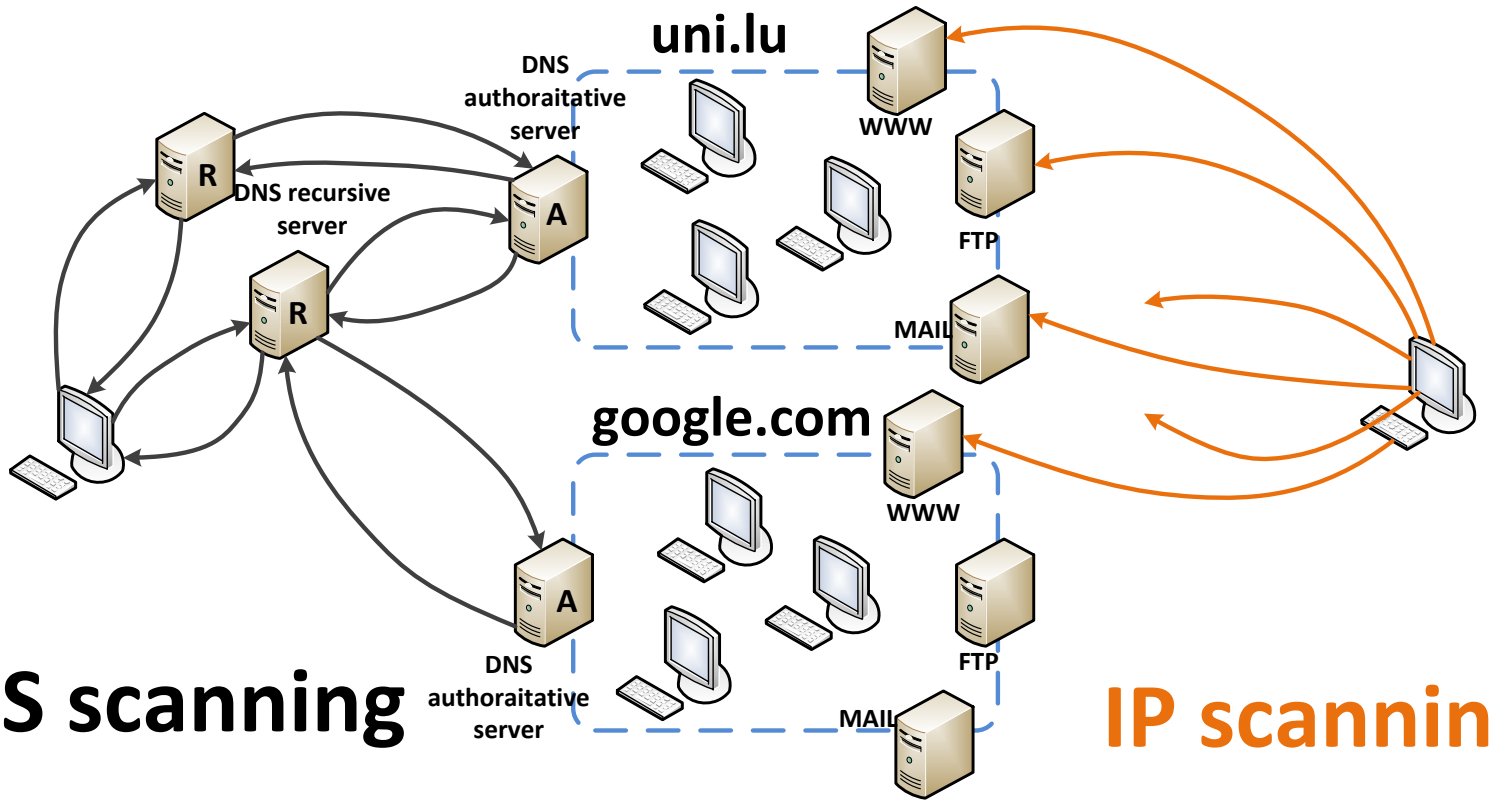
Why DNS brute forcing ?

DNS scanning:

- Test DNS names by sending requests to a DNS recursive server
- Network mapping: discover all machines of a domain
  - penetration testing
  - network security assessment (prevention)
  - recon stage to craft attack (find accessible services)
  - use by worms to spread themselves

Alternative to IP scanning

- Provided services (ftp, www, etc.) $\neq$ network interfaces
- Machines are not probed directly (DNS requests)
- Can be enhanced by using multiple open recursive servers
- Reduce the search space (particularly in IPv6)



SNT

Current Approaches

- How names are defined?
  - by human and easy to remember → pc1, pc2, atlanta, boston, etc.
  - to reflect the provided service → www, ftp, ssh, etc.
- → same names often used → scan the most popular names
- dictionary based tools
  - DNSenum: 266 930 names by default
  - fierce: 1 895 names by default
- tool relying on natural language
  - SDBF: domain name generator (domain names features, Markov chain model)

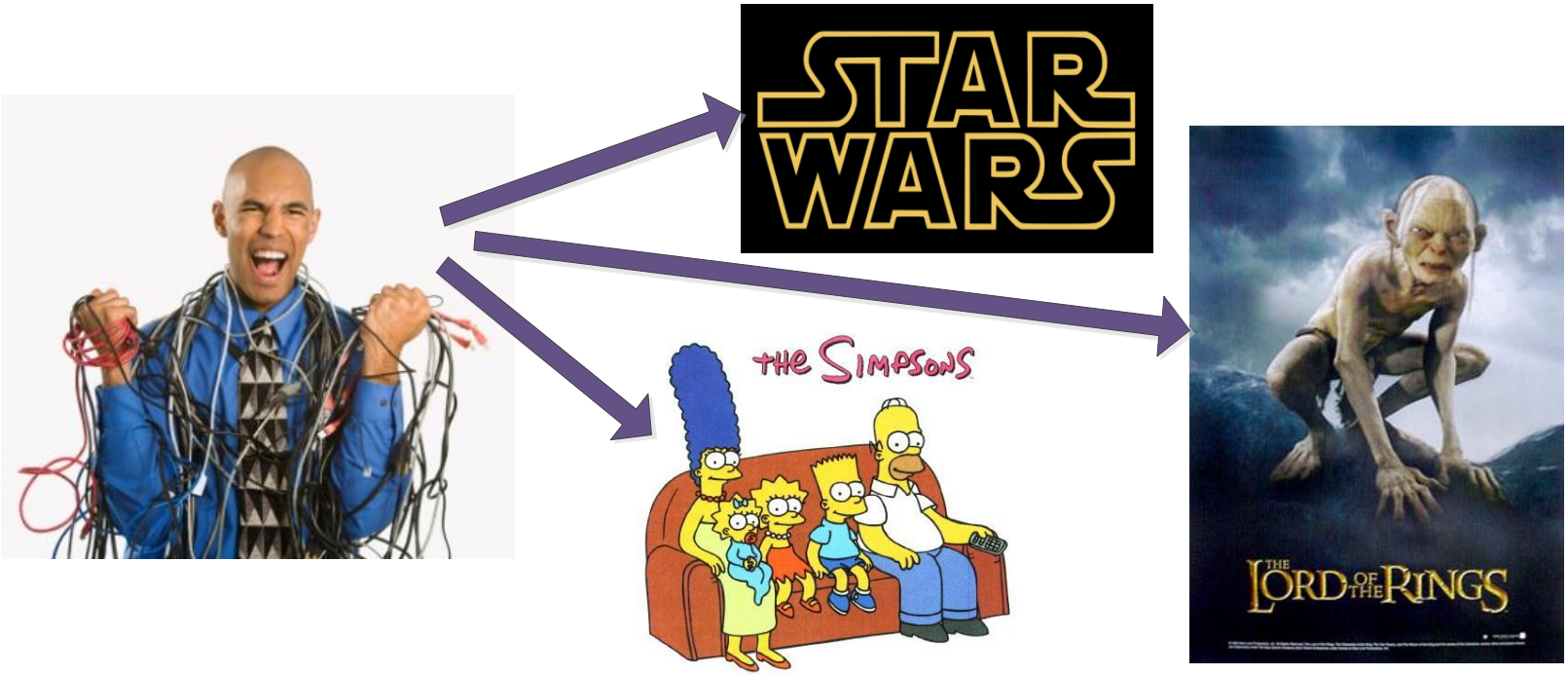
SMT

Outline

Names definition

Domain names are given by human:

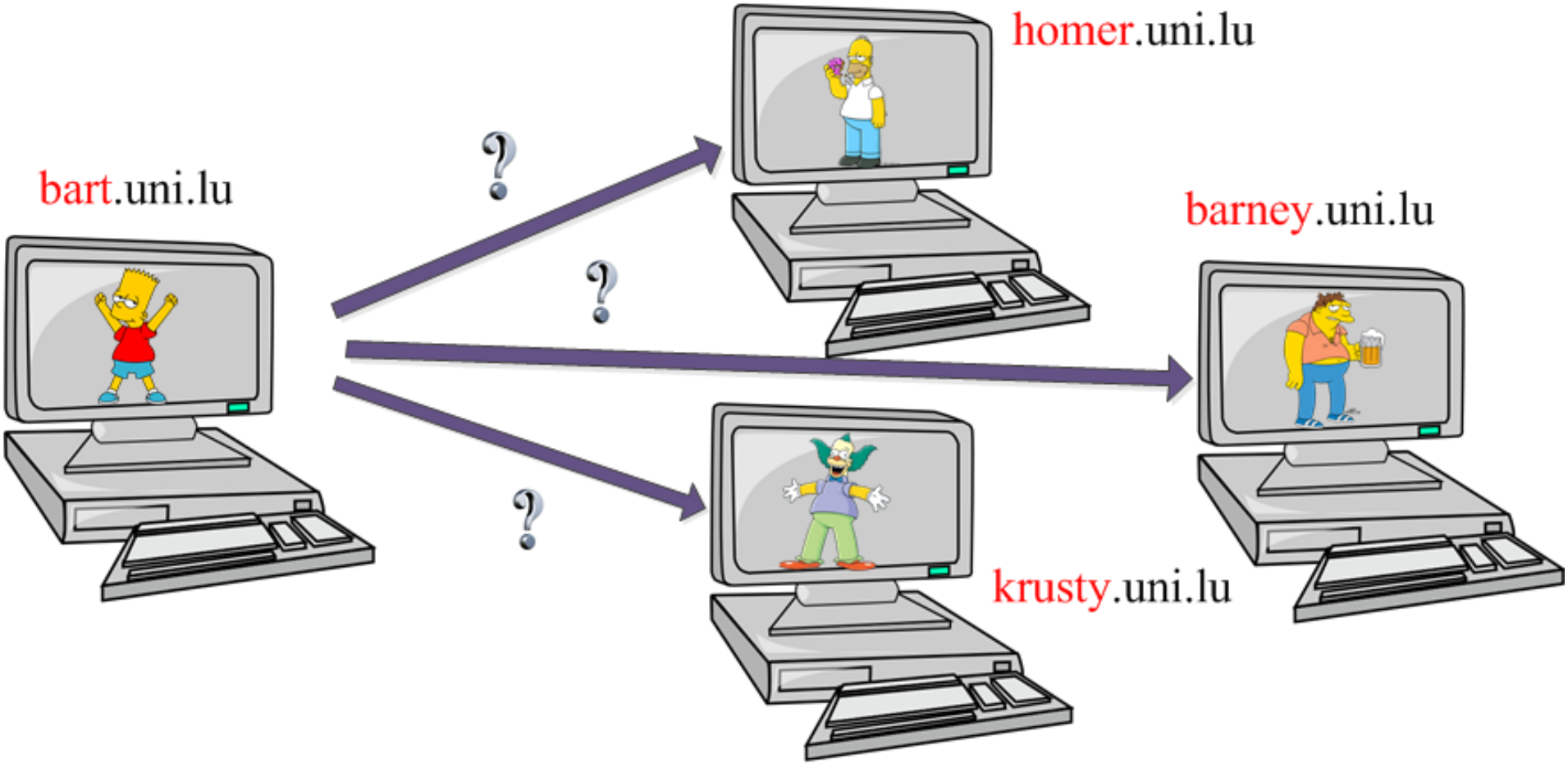
- Reflect a provided service (ftp, www, etc.)
- Follow numerical patterns (ftp1, ftp2)
- Share a common semantic field



SNT

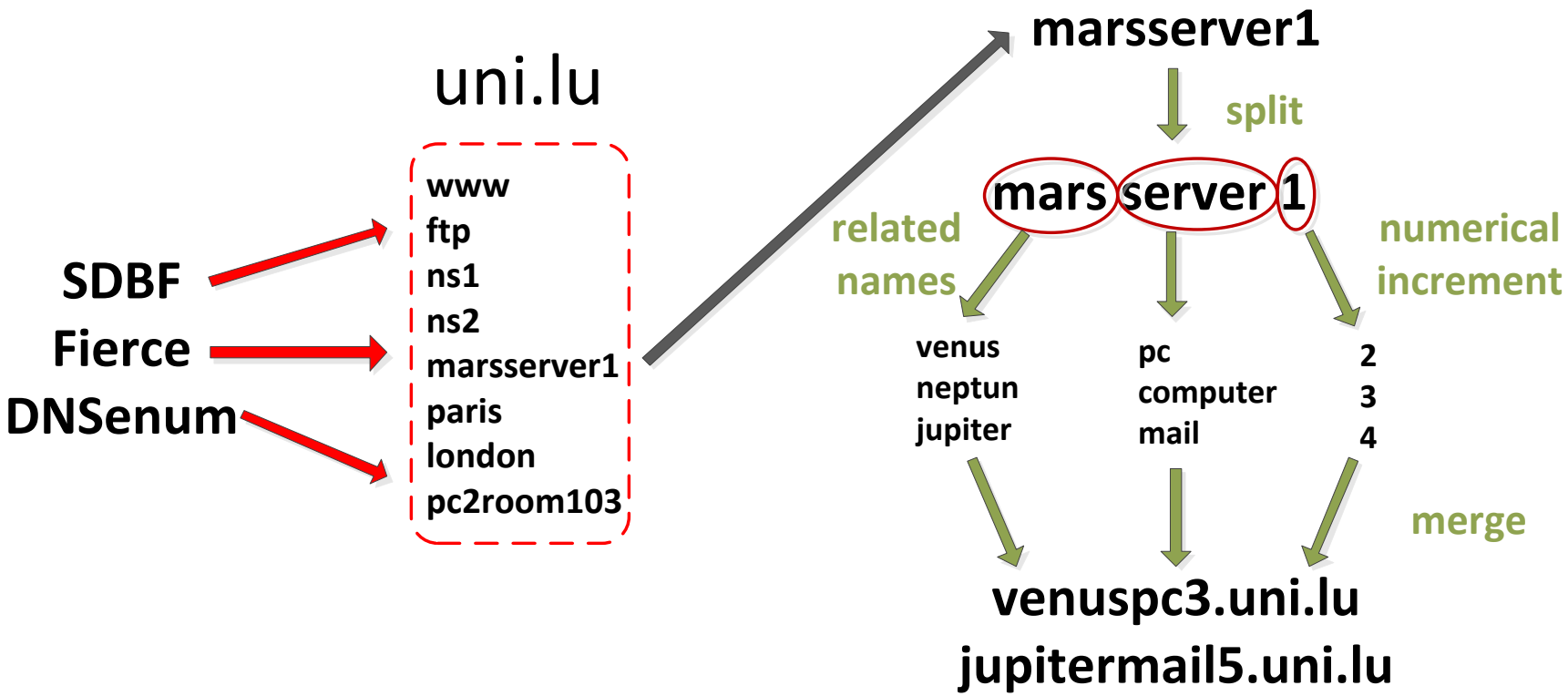
Names discovery

Given a subdomain ⇒ generate new subdomains:

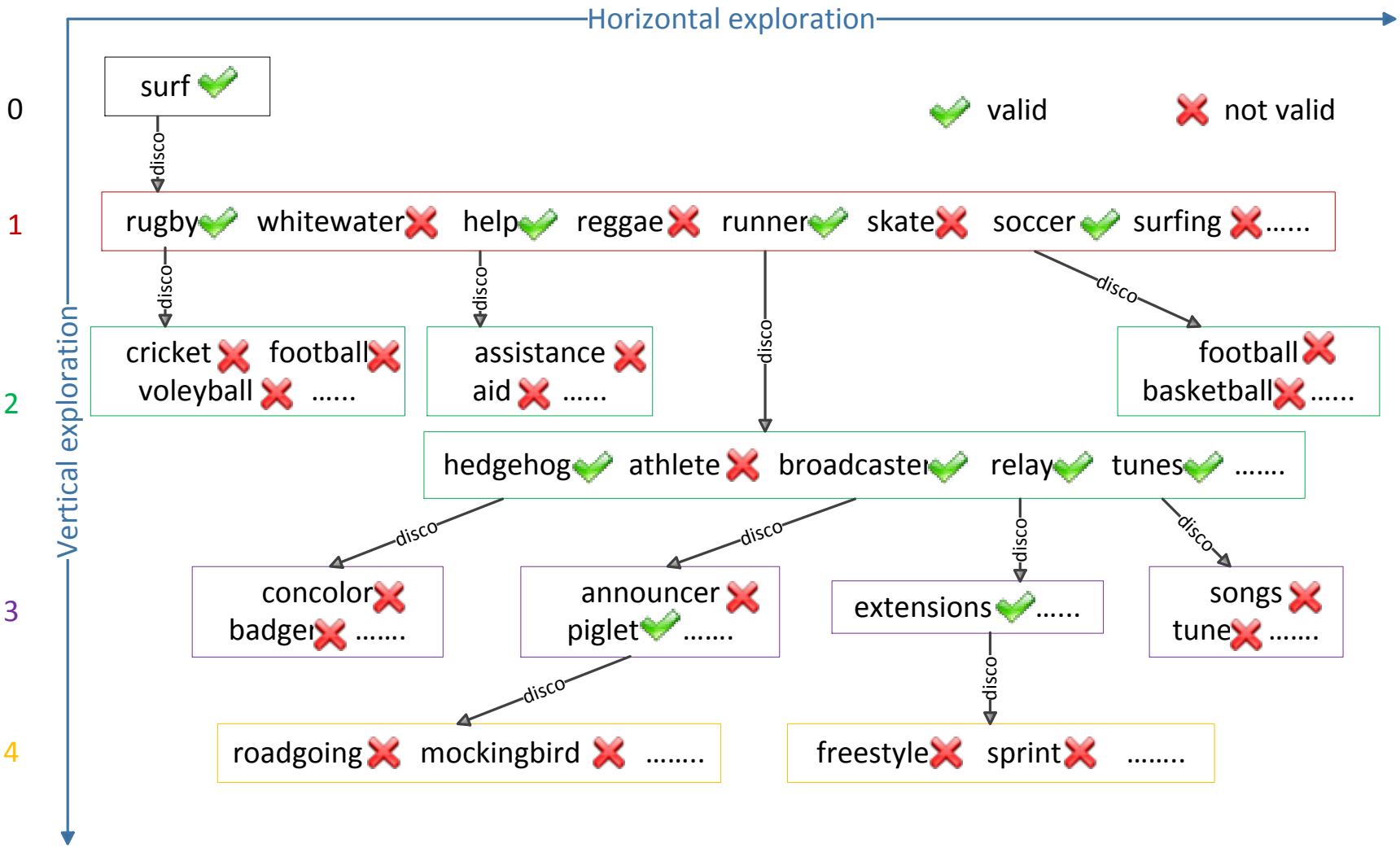


Tool overview

- Generation of semantically close names
  - name split in words
  - generate similar words using Disco
- Enumeration of numbers



Semantic exploration: Disco



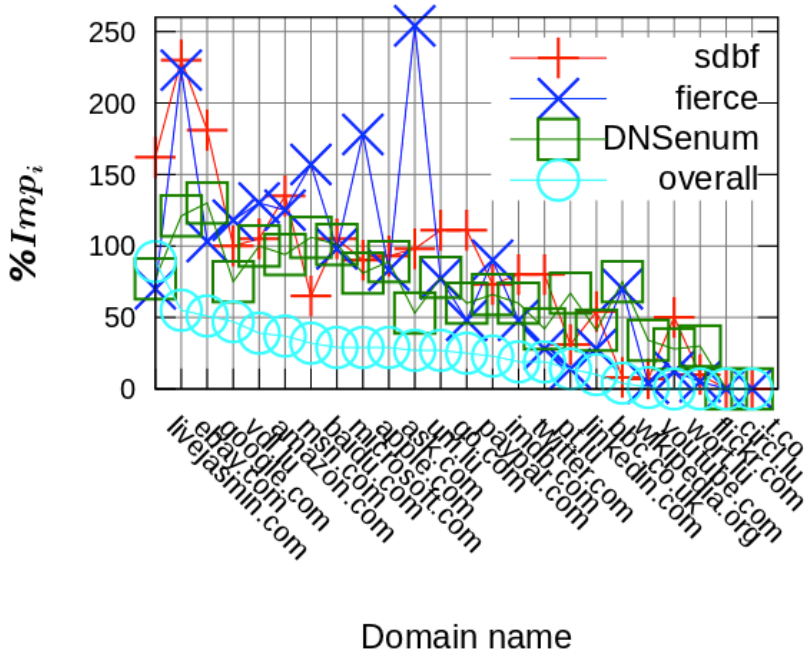
SMT

Outline

Method and Results

- 24 popular domains
- 3 tools: DNSenum, Fierce, SDBF $\Rightarrow$ initial list of subdomains

$$\%Imp_i = \frac{|New_i|}{|Init_i|}, i \in \{SDBF, DNSenum, Fierce, overall\}$$

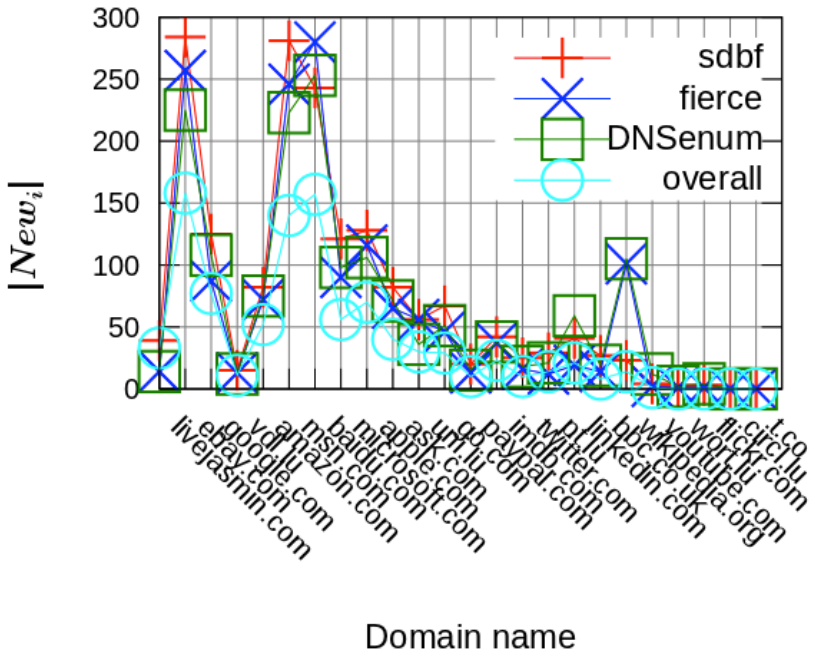


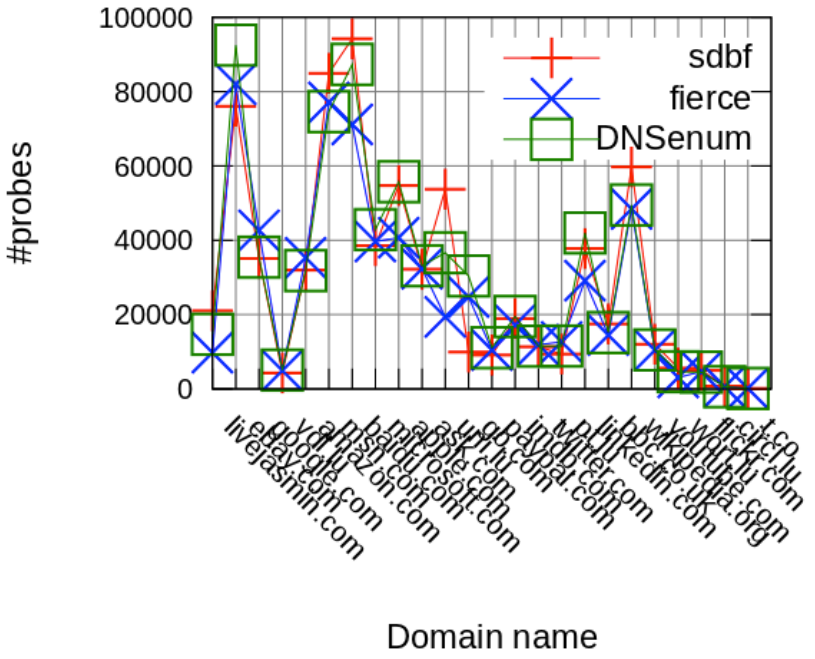
Detailed Results

	SDBF			Fierce			DNSenum			Overall		
Domains	$\|Init\|$	$\|New\|$	$\%Imp$	$\|Init\|$	$\|New\|$	$\%Imp$	$\|Init\|$	$\|New\|$	$\%Imp$	$\|Init\|$	$\|New\|$	$\%Imp$
livejasmin.com	24	39	162	20	14	70	18	14	77	37	33	89
ebay.com	123	284	230	115	257	223	185	225	121	284	158	55
google.com	69	125	181	84	87	103	83	108	130	149	77	51
vdl.lu	15	15	100	11	13	118	16	12	75	23	11	47
amazon.com	78	82	105	55	72	130	75	75	100	132	52	39
msn.com	207	281	135	196	246	125	236	223	94	372	140	37
baidu.com	369	243	65	178	280	157	238	253	106	478	157	32
microsoft.com	115	121	105	91	90	98	97	98	101	189	56	29
apple.com	141	128	90	65	116	178	130	106	81	241	70	29
ask.com	88	82	93	78	65	83	79	71	89	135	40	29
all domains	2057	1739	84	1520	1558	102	1788	1565	87	3170	954	30

- From 84% to 102% of newly discovered names
- Up to 230% of improvement
- Complementarity $\Rightarrow$ 30 % overall improvement

Overhead





- Average of 40,000 probes per domain // SDBF & DNSenum: 260,000 (6 times less)

SMT

Outline

SNT

Conclusion

- New methods to brute-force DNS:
  - semantic relatedness
  - incremental techniques
- Results:
  - able to generate valid names...
  - ... mainly not present in well used dictionaries → complementarity
  - low overhead
- Future works:
  - use other databases
  - improve semantic relatedness metric

Semantic Exploration of DNS

Samuel Marchal, Jérôme François, Cynthia Wagner and Thomas Engel

UNIVERSITÉ DU LUXEMBOURG

SnT

NETLAB