AFRL-RI-RS-TR-2012-002

GROUNDING UNDNS

UNIVERSITY OF MARYLAND

JANUARY 2012

FINALTECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. This report is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RI-RS-TR-2012-002 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR:

/s/

ROBERT L. KAMINSKI
Work Unit Manager

/s/

WARREN H. DEBANY JR., Technical Advisor
Information Exploitation and Operations Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
JAN 2012	Final Technical Report	JUL 2010 – SEP 2011

4. TITLE AND SUBTITLE

GROUNDING UNDNS

5a. CONTRACT NUMBER
FA8750-10-2-0194

5b. GRANT NUMBER
N/A

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Neil Spring

5d. PROJECT NUMBER
BYU1

5e. TASK NUMBER
MA

5f. WORK UNIT NUMBER
RY

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Maryland
Office of Research
Administration & Advancement
College Park MD 20742-5100

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory/RIG
525 Brooks Road
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RI

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-RI-RS-TR-2012-002

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This effort applied the DNS names, IP addresses, and IP alias relationships collected as part of the Discarte and RadarGun projects to instrument the Undns rules, developing robust metrics for completeness and location imprecision. Undns, a DNS-based router location inference sy7stem, comprises three parts" an engine for partial regular expression matching customized for DNS names, a set of rules that comprise over 12,000 lines spread across approximately 270 autonomous systems, and an interface to scripting languages (Perl and Ruby) to permit on-line use. Armed with these metrics, the Undns ruleset is better revised, vestigial rules removed or demoted for maintenance, and redundant locations distinguished.

15. SUBJECT TERMS
Domain Name Service (DNS), Router Location Interference System

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			ROBERT KAMINSKI
U	U	U	UU	16	19b. TELEPHONE NUMBER (*Include area code*) N/A

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

TABLE OF CONTENTS

Section **Page**

Grounding Undns Final Report

1 Introduction and Background

As part of Rocketfuel, the Principal Investigator (PI) created Undns [8], a DNS-based router location inference system that has acted as a foundation, or a reference, for several research projects that require geographic annotation [6 ,10 ,4 ,2 ,9 ,3]. Undns comprises three parts: an engine for partial regular expression matching customized for DNS names, a set of rules that comprise over 12,000 lines spread across approximately 270 autonomous systems,[1] and an interface to scripting languages (Perl and Ruby) to permit on-line use. Current Undns software and rules are available at https://subversion.umiacs.umd.edu/undns/trunk/.

Undns is successful enough that others have contributed their updates to the rules. The rulesets have been extended by Freedman et al. [2] in January 2006, and by Madhyastha et al. [5] in October 2006. Ying Zhang et al. [11] also revised the rules for a 2009 paper. The PI credits the success of Undns to two features: the immediacy of DNS-based positioning—that it does not require active measurement to triangulate positions—and the ease with which researchers can incorporate the code into their projects.

However, there are limitations, both inherent to Undns and to its evolution. Undns places trust in DNS names—that they are assigned correctly to interfaces that do not move. It also trusts the inference of the underlying locations by abbreviation—that the identifier clearly specifies the city, despite the multiple cities named Vancouver or Paris, for example. These potential errors are either inherent (relying on ISP administrators) or by experimenters (relying on accurate inference of the location). As a practical matter, the potential that poor rules would be written means that it is an unsolved challenge to ensure that new rules are bug-free and worthy of being used by others.

In this project, the PI proposed to engage in two activities: instrumentation of rule applicability, and the instrumentation of location accuracy. Time degrades the utility of Undns rules. ISPs merge, expand, incorporate other networks, etc., creating "new" locations unknown to the rules. Similarly, rules that recognize (city-scale) locations assert precision that is not practically present. Concretely, those ISPs that use airport-code identities for locations may use IAD to refer to Sterling, VA (very near Dulles), Ashburn (somewhat near Dulles but more developed with network), or perhaps even DC.

1 The largest ISP ruleset is for Comcast at 757 lines.

```
9264 ¥.ascc¥.net$ {
    /* Academia Sinica Computing Center */
    ¥.taipeigigapop¥.ascc¥.net$ loc="Taipei, Taiwan";
    ¥.tw¥.ascc¥.net$ loc="Taipei, Taiwan";
    ¥.hk¥.ascc¥.net$ loc="HongKong";
}
```

Figure 1: Simple example of an Undns ruleset for AS 9264 using the DNS suffix ascc.net.

```
9264 ¥.ascc¥.net$ {
    /* Academia Sinica Computing Center */
    ¥.taipeigigapop¥.ascc¥.net$ loc="Taipei, Taiwan";
                            /* [dc: 4 ip, 1 rtr, consistent ] */
    ¥.tw¥.ascc¥.net$ loc="Taipei, Taiwan";
                            /* [dc: 12 ip, 3 rtr, consistent ] */
    ¥.hk¥.ascc¥.net$ loc="HongKong";
                            /* [dc: 1 ip, 1 rtr, consistent ] */
    /* Unmatched [dc: 20 ip, 18 rtr] */
}
```

Figure 2: **Proposed annotated ruleset, including with the frequency with which a rule was matched, the number of apparently distinct routers using those names, and whether aliases of that name have identical (consistent) location. Names may match the top-level prefix but lack a matched location. The format or detail of the annotations would be developed as part of this proposal.**

Rule development for Undns has largely followed an ad hoc procedure: compute locations for as many addresses as possible, sort and browse names that lack location mappings looking for possibilities for significant new rules, and write rules based on inferred location, iterating until tired. This procedure has two main failings: it is difficult to determine the most salient rules to write, and it is difficult to realize when ambiguous locations are problematic. Notable ambiguous locations include “Springfield” (typically Massachusetts or Virginia, not Illinois) and London (possibly Canada).

The PI proposed to apply the DNS names, IP addresses, and IP alias relationships collected as part of the Discarte[7] and RadarGun [1] projects to instrument the Undns rules, developing robust metrics for completeness and location imprecision. Armed with these metrics, the Undns ruleset can be better revised, vestigial rules removed or demoted for maintenance, and apparently redundant locations distinguished. (For example, AT&T’s various locations in Chicago, “chcil”, “chgil”, and “cgcil”, if seen frequently enough, should be given distinct geographic coordinates.)

The intent was to generate in-line instrumentation sketched in Figures 1and 2. The PI noted challenges in providing this in-line instrumentation having to do with common use of single

tables, e.g., the list of airport codes is included in several conventions. The PI did not appreciate the difficulty of discovering good annotation locations through the lex and yacc parser, nor did the PI recognize that matching rules and doubting rules are separate operations. As a result, the PI produced a representation of the error-analysis annotations, but stopped short of feeding back those annotations into the original rules.

The proposed deliverable from this project was to be a revised Undns software package consisting of two elements: first, a statistics engine that tracks observations of addresses bearing matching hostnames and observations of aliases having consistent or inconsistent location, and second, a revised ruleset annotated by how each rule applies to Discarte data. The first is entirely complete: every rule has statistics about whether it is matched and whether it is universally consistent. The second task is not, with the recognition that the main source of error is what is missing from the rules, annotating stale rules and embedding that a rule might be dubious in the ruleset itself seemed a mistake. Keeping this information separate, as a product of a separate analysis that would feed back (manually) into the rules was more effective.

2 Annotation Design

Although the counters employed here are not substantially novel, their structure exposes some partially-addressed challenges in instrumenting the ruleset.

In Undns, each hostname matches first against a "pre-filter," intended to guard against unnecessarily matching complex regular expressions. After the pre-filter is a "convention", a regular expression that includes parenthesized expressions to be extracted from the name. If the convention matches, the extracted fragments are matched against "key-value" pairs. These pairs might match "ashbva" to "Ashburn, VA". As a postprocessing step, a "Location Table" maps the city name to a lat/long coordinate. The detail of this location table is only for the broader city, without knowledge of the buildings that house network equipment.

To annotate, the PI added two counters to each "convention", "key-value" bindings (as a whole), and each "key-value". These two counters are *matches* and *doubts*. If a convention matches, the match counter is incremented. If there's a reason to doubt the match (i.e., the location is inconsistent with the location of aliases as in this analysis), the doubt counter is incremented. The same logic applies to each individual key-value mapping. The key-value group (as a whole) increments the doubt counter whenever the convention matched but no key value was found. This occurred especially often for Cox communications, which uses a somewhat opaque naming convention, in which it's clear what part of the name represents a location but it's less clear what location that happens to be.

The match counter is automatically incremented; the doubt counter is explicitly set by the analysis, except when a key-value is missing. Concretely, the PI added two methods,

doubt_location to note a possible mistake in the location and dump_parsed_conventions_to_file to dump the counters to a file.

In the alias-based analysis, the match and doubt counters are not comparable: there may be many more doubts than matches if an alias appears incorrect. The PI expected the high doubt count to be a clearer indication of a mistake. However, the potential for many doubts on few matches means that the "worst" rules in terms of doubt-to-match ratio are rarely matched.

3 Classified Mistakes

Each "mistake" is an instance of a router that has aliases in "different" locations. The aliases may be false, but more likely, the names are incorrect or incorrectly decoded.[2] The PI classified the errors into four categories: the product of same or different asn, and below or above 1 ms geographic distance. The intuition is that the errors that cause these four classes of mistake are likely to be different, and that these four classes of error are reasonably easy to recognize automatically.

Table 1: Mistakes where alias pairs have different locations, classified by AS similarity and distance (1ms threshold).

	same AS	different AS
small / short / near	401	261
large / long / distant	824	197

3.1 Large Distance, Same AS

The dominant category of mistake is those where both names belong to the same naming convention, and simply have different locations. An example is below:

```
 Columbia, SC != Atlanta, GA (66.35.174.90, 66.35.174.202) (pos3-
0.clmascea14w.cr.deltacom.net, gig2-0.atlngapk24w.xr.deltacom.net)
 Columbia, SC != Atlanta, GA (66.35.174.90, 66.35.174.105) (pos3-
0.clmascea14w.cr.deltacom.net, pos5-0.atlngapk24w.cr.deltacom.net)
 Columbia, SC != Atlanta, GA (66.35.174.90, 66.35.174.5) (pos3-
0.clmascea14w.cr.deltacom.net, pos3-1.atlngapk24w.cr.deltacom.net)
 Columbia, SC != Atlanta, GA (66.35.174.90, 66.35.174.25) (pos3-
0.clmascea14w.cr.deltacom.net, pos1-0.atlngapk24w.cr.deltacom.net)
```

Only these four mismatches appeared for deltacom.net, and there were other addresses with Columbia names (clmasc). This single address is likely to be in Atlanta instead of its classified

2 I have asked CAIDA researchers for recent output of their MIDAR tool to try to get a current, large, and perhaps more accurate set of aliases; they have promised one after their next run.

Columbia.

A second example suggests an error in alias resolution for cenic.net. This research network does not often respond to alias probing, so it is likely that these aliases were inferred by Discarte's analysis incorrectly, combining Los Angeles, Sunnyvale, and Oakland routers. An excerpt of the 280 pairwise mismatches this caused follows.

```
Sunnyvale, CA != LosAngeles, CA (137.164.22.30, 137.164.22.20) (dc-svl-dc1--
oak-dc1-10ge.cenic.net, dc-lax-dc1-lax-dc2-ge--3.cenic.net) same-asn 0.0024s
Sunnyvale, CA != LosAngeles, CA (137.164.22.30, 137.164.22.24) (dc-svl-dc1--
oak-dc1-10ge.cenic.net, dc-lax-dc1--slo-dc2-pos.cenic.net) same-asn 0.0024s
LosAngeles, CA != Oakland, CA (137.164.22.228, 137.164.40.81) (dc-lax-dc1--
riv-dc1-pos.cenic.net, dc-oak-dc2--csusanfran-egm.cenic.net) same-asn 0.0026s
LosAngeles, CA != Oakland, CA (137.164.22.228, 137.164.32.16) (dc-lax-dc1--
riv-dc1-pos.cenic.net, dc-oak-dc2--contracostacoe-ds3.cenic.net) same-asn
0.0026s
```

A third case appears be an error of locating subnet gateways. ISPs seem to assign a location-bearing name to all addresses on a subnet, including the ".1" likely to be the subnet gateway, even though the gateway router is not necessarily in the suburb with the clients. An example follows in which the addresses having location inferred to be in Monterey (mty) or Mexico City (mx) are matched as aliases. (Four of 18 matches are shown.) Since each of the failed addresses is a ".1", it seems that this tag is only an indication of location for the clients of this subnet. The router itself is more likely to be in the larger city. A similar naming scheme accounts for 50 mismatches between Fort Worth and Denton, TX in charter.com.

```
MexicoCity, Mexico != Monterrey, Mexico (200.39.118.1, 200.56.228.1) (ip-200-
39-118-1-mx.marcatel.net.mx, ip-200-56-228-1-mty.marcatel.net.mx) same-asn
0.0034s
MexicoCity, Mexico != Monterrey, Mexico (200.39.115.1, 200.56.228.1) (ip-200-
39-115-1-mx.marcatel.net.mx, ip-200-56-228-1-mty.marcatel.net.mx) same-asn
0.0034s
MexicoCity, Mexico != Monterrey, Mexico (200.53.37.1, 200.56.228.1) (ip-200-
53-37-1-mx.marcatel.net.mx, ip-200-56-228-1-mty.marcatel.net.mx) same-asn
0.0034s
MexicoCity, Mexico != Monterrey, Mexico (200.39.119.1, 200.56.228.1) (ip-200-
39-119-1-mx.marcatel.net.mx, ip-200-56-228-1-mty.marcatel.net.mx) same-asn
0.0034s
```

In summary, in this category, the Undns rules appear good. The apparent bulk of the mismatches seems caused by relatively few faulty aliases which appear as many incorrect locations. The bulk of the original errors seem to be from names that are incorrect or stale, but are otherwise correctly decoded. A practice of naming the gateway based on the subnet suggests that such IP-address-based names and the .1 addresses should be discounted if used in any voting-type

scheme for assigning a location to a router based on the majority of locations.

3.2 Small distance, Same AS

This category is largely the same in profile as the large distance variant. (197 of the errors were from the cenic.net false alias.) There are more of the ".1" third category above, since the distances are typically small.

Added to these categories are some that distinctions in name only, often from the DC suburbs or Palo Alto. A particularly easy-to-fix version was that the lower case "sterling" was mapped to "Sterling, VA", while the upper case was mapped to "Washington, DC" for savvis.net:

```
Sterling, VA != Washington, DC (216.109.66.33, 216.33.98.153) (bhr1-g12-
0.sterling2dc3.savvis.net, bhr1-g9-2.Sterling2dc3.savvis.net) same-asn 0.0002s
Washington, DC != Sterling, VA (216.33.96.217, 206.24.227.42) (bhr1-g8-
0.Sterling2dc3.savvis.net, bhr1-ge-3-1.sterling2dc3.savvis.net) same-asn
0.0002s
```

Similar errors noted that "LosAngeles, CA" was not "Los Angeles, CA". Two-word country names such as New Zealand had to be combined to one word for ease of parsing, but two-word city names were okay. This led to confusion and inconsistent cities.

3.3 Small distance, Different AS

Routers may have IP addresses from various autonomous systems, since addresses are part of networks and networks are assigned to these organizations. In order to connect two ISPs, at least one router must have an address from both ISPs. It is each ISP's responsibility to populate the hostnames for these addresses, so the router will then have names matching distinct conventions.

The different AS class is more likely to expose small differences between how we have decoded locations at peering points. For example, 222 of the 261 errors involved Ashburn, VA, typically because some ISPs are specific, others use the airport code for Dulles, IAD.

```
Ashburn, VA != Washington, DC (63.216.0.93, 154.54.12.110) (fe7-
6.cr01.ash01.pccwbtn.net, btn.iad01.atlas.cogentco.com) asn-mismatch 0.0002s
```

Of the remaining, there are a few more with the two-word city names, and some that show an error in generic-vs-specific naming. A default rule maps unmatched Israel ".il" to Tel Aviv, but a specific name matches Haifa. The best approach to solve this problem is unclear, though it does appear in the rules: although the .il rule matched 57 times, it accumulated 57 doubts, i.e., perhaps it did not return the right answer even once.

```
TelAviv, Israel != Haifa, Israel (194.90.151.1, 212.143.8.7)
```

(SQLgate.netvision.net.il, gi1-1.srvc1.hfa.nv.net.il) asn-mismatch 0.0003s
TelAviv, Israel != Haifa, Israel (194.90.151.1, 207.232.53.1)
(SQLgate.netvision.net.il, vl610.srvc.hfa.netvision.net.il) asn-mismatch
0.0003s

A similar error affects a Romanian ISP, where the default binds to Bucharest but the specific notices Ploiesti. However, since the Ploiesti address is a .1, and thus of lower confidence as described above, perhaps the generic rule is correct.

Ploiesti, Romania != Bucharest, Romania (81.196.221.1, 213.154.113.131) (host-81-196-221-1.ploiesti.rdsnet.ro, 213-154-113-131.rdsnet.ro) asn-mismatch
0.0003s

3.4 Large distance, Different AS

The final category of the four are large distance, different AS. This is where the host names belong to the naming conventions of different ISPs, but the IP addresses are aliases.

There are again cases of apparently bad names (one IP address uses Anaheim as a location even though all its aliases appear to be in Sydney):

Anaheim, CA != Sydney, Australia (64.200.142.174, 203.111.1.37) (anhmca1wct1-powertel-atm.wcg.net, ge-1-1-36.syd-core-p-01.powertel.net.au) asn-mismatch
0.0604s
Anaheim, CA != Sydney, Australia (64.200.142.174, 202.92.64.53) (anhmca1wct1-powertel-atm.wcg.net, syd-core-p-01-ge2033.powertel.net.au) asn-mismatch
0.0604s
Anaheim, CA != Sydney, Australia (64.200.142.174, 202.92.64.154) (anhmca1wct1-powertel-atm.wcg.net, syd-core-p-01-ge216.powertel.net.au) asn-mismatch
0.0604s

There are cases like those above in which generics (by country code or by ISP domain) match a different location from another ISPs specifics. In the excerpt below, first an ISP in Brazil defaults to the capital; second a Taiwanese ISP has to peer in San Jose but the rules do not recognize the peering

SaoPaulo, Brazil != RioDeJaneiro, Brazil (201.38.38.49, 201.45.200.174) (eth0-atm.marinter.com.br, marinter-S11-1-2-acc17.rjo.embratel.net.br) asn-mismatch
0.0018s
SaoPaulo, Brazil != RioDeJaneiro, Brazil (201.38.38.49, 201.45.200.170) (eth0-atm.marinter.com.br, marinter-S11-0-3-acc17.rjo.embratel.net.br) asn-mismatch
0.0018s
SanJose, CA != Taipei, Taiwan (154.54.13.18, 198.32.176.21)
(hinet.sjc04.atlas.cogentco.com, ge.pa-c12r11.USA-PA.router.hinet.net) asn-mismatch 0.0521s
Taipei, Taiwan != SanJose, CA (211.72.108.145, 154.54.11.130) (pa-c12r11.USA-PAIX.router.hinet.net, hinet.sjc04.atlas.cogentco.com) asn-mismatch 0.0521s

Finally, the rest of the faulty ".il" locations are in this category, in which an Israeli ISP peers using addresses that lack specific location information, connecting to a router with specific location information.

```
TelAviv, Israel != Palermo, Italy (212.199.18.130, 195.22.197.50)
(212.199.18.130.forward.012.net.il, customer-side-goldenlines-5-
il.pal6.pal.seabone.net) asn-mismatch 0.0101s
TelAviv, Israel != London, UnitedKingdom (212.199.73.69, 63.218.13.42) (pt-
212.199.73.69.static.012.net.il, goldenlines.pos4-7.ar03.ldn01.pccwbtn.net)
asn-mismatch 0.0178s
```

4 The worst rules

The ten "worst" rules are as follows:

```
    incomplete keys: ../keys/cox.22773:13 - matches: 4146, doubts: 3080
    incomplete keys: ../keys/embratel.4230:8 - matches: 667, doubts: 1238
    incomplete keys: ../keys/blueyonder.co.uk.5462:7 - matches: 706, doubts:
1130
  2152, ^(dc-|hpr-|inet-)([a-z]{3,4})-.*¥.cenic¥.net, ../keys/cenic.2152:5 -
matches: 1577, doubts: 957
    incomplete keys: ../keys/swbell.7132:63 - matches: 2774, doubts: 467
    incomplete keys: ../keys/blueyonder.co.uk.5462:4 - matches: 6, doubts: 422
    incomplete keys: ../keys/bteu.5400:22 - matches: 71, doubts: 396
    incomplete keys: ../keys/charter.many:122 - matches: 862, doubts: 392
      svl->Sunnyvale, CA ../keys/cenic.keys:12 - matches: 42, doubts: 367
    incomplete keys: ../keys/cw.3561:114 - matches: 150, doubts: 363
```

Each "incomplete keys" rule represents a match of a convention that lacks a key-value pair. A bit of investigation can likely fill these missing entries in, though with this information it is clear which conventions need attention. The fourth rule (cenic) represents the faulty alias problem that dominated the mismatches above. Note that incompleteness will not produce an inconsistency, the worst of the inconsistencies is dwarfed by the missing keys for cox, embratel, blueyonder, charter, and cw.

Cox will take some understanding; their convention seems to include a pair of two-character location tags, ".br.br" is the most common missed location. This appears (by name) to be Baton Rouge, Louisiana, but it is unconvincing. More convincing would be ".ok.ok" as Oklahoma City, Oklahoma. Perhaps the rule is that the city is a two-character combination and the main point of presence is another two-character combination, so ".lf.br" is Lafayette, near Baton Rouge.

The ten worst rules, by doubt-to-match ratio are as follows:

```
    incomplete keys: ../keys/blueyonder.co.uk.5462:4 - matches: 6, doubts: 422
    incomplete keys: ../keys/forthnet.1241:4 - matches: 1, doubts: 42
    incomplete keys: ../keys/ctnet.co.jp.7670:5 - matches: 2, doubts: 78
```

```
  ashva->Ashburn, VA ../keys/aleron.4200:8 - matches: 4, doubts: 108
incomplete keys: ../keys/home.ne.jp.9824:10 - matches: 7, doubts: 170
incomplete keys: ../keys/retevision.es.8761:17 - matches: 1, doubts: 20
incomplete keys: ../keys/impsat.19583:7 - matches: 6, doubts: 98
incomplete keys: ../keys/telstra.1221:21 - matches: 4, doubts: 50
incomplete keys: ../keys/demon.2529:11 - matches: 7, doubts: 82
incomplete keys: ../keys/rr.many:21 - matches: 11, doubts: 126
```

The Ashburn error for aleron (the fourth line) is as described above: an alias with a router tagged as "IAD" in another ISP.

5 Summary of Conclusions

The errors in Undns, in decreasing order of importance are:

Lacking Completeness

: Even for the naming conventions where we understand where a location-identifying-fragment is supposed to be, relatively many locations are missing. 3,000 IP addresses in Cox alone could have been given a location if only the fragments in known conventions were better understood. There are another 10,000 addresses in RoadRunner (rr.com) discovered by Discarte that appear to have location information, but are not given a location by the rules, which would be a similar problem except that RoadRunner is unique in that it uses different AS numbers for different regions and this conflates with the name-based rules.

These large ISPs are perhaps easily fixed individually, but the long tail of smaller ISPs is likely inescapable.

Failed or Dubious Alias Resolution

: Any faulty alias creates the appearance of massive error in locating routers. The mismatch statistics for the "same ISP" categories above were dominated by what appears to be a faulty set of aliases in cenic.net; other apparently faulty aliases also inflated the mismatches.

Although seeing different locations for IP addresses within an alias could be evidence that the alias is wrong, we have also seen cases where it is evidence that a name is wrong. Analysis could distinguish the exceptional cases, such as cenic.net above as a bad alias, or Columbia above as a bad name, but may have trouble deciding which is incorrect in other scenarios, such as the Monterrey vs. Mexico City error above.

Inconsistency across ISPs

: This comparison exposed mistakes in how cities are named (spaces or not) and in how DC and bay area suburbs are given location names by different ISPs.

Such errors may increase as suburbs develop, but being able to compare across ISPs as done here may help contain the inconsitency.

Equal weight to Generics and .1's

: Although tempting to put all .tw names in Taiwan and all .il names in Israel, those inferred locations will not be correct for peering points, and may not be specific enough for countries as they develop. Similarly, the .1 addresses of dynamically-assigned subnets may be on routers far from the service area. Neither generic names, or the served locations of .1 addresses should weigh as heavily in deciding on the location of a router. The interface of "what is the location for this hostname" does not include such weighting information.

To correct this would require a new interface that would either take a set of names and IP addresses to return the most likely location despite ambiguity, or continue to take a single name and return some confidence level with each location.

6 References

[1] Adam Bender, Rob Sherwood, and Neil Spring. Fixing Ally's growing pains with velocity modeling. In *Proceedings of the ACM SIGCOMM Internet Measurement Conference (IMC)*, Vouliagmeni, Greece, October 2008.

[2] Michael J. Freedman, Mythili Vutukuru, Nick Feamster, and Hari Balakrishnan. Geographic locality of IP prefixes. In *Proceedings of the ACM SIGCOMM Internet Measurement Conference (IMC)*, Berkeley, CA, October 2005.

[3] Cheng Huang, Angela Wang, Jin Li, and Keith W. Ross. Measuring and evaluating large-scale CDNs. In *Proceedings of the ACM SIGCOMM Internet Measurement Conference (IMC)*, Vouliagmeni, Greece, October 2008.

[4] Ethan Katz-Bassett, John P. John, Arvind Krishnamurthy, David Wetherall, Thomas Anderson, and Yatin Chawathe. Towards IP geolocation using delay and topology measurements. In *Proceedings of the ACM SIGCOMM Internet Measurement Conference (IMC)*, pages 71–84, Rio de Janeiro, Brazil, October 2006.

[5] Harsha V. Madhyastha, Tomas Isdal, Michael Piatek, Colin Dixon, Thomas Anderson, Aravind Krishnamurthy, and Arun Venkataramani. iPlane: An information plane for distributed services. In *Proceedings of the Symposium on Operating Systems Design and Implementation (OSDI)*, Seattle, WA, November 2006.

[6] Ratul Mahajan, Ming Zhang, Lindsey Poole, and Vivek Pai. Uncovering performance differences among backbone ISPs with Netdiff. In *Proceedings of the ACM/USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, pages 205–218, San Francisco, CA, April 2008.

[7] Rob Sherwood, Adam Bender, and Neil Spring. Discarte: A disjunctive internet cartographer. In *Proceedings of the ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication (SIGCOMM)*, Seattle, WA, August 2008.

[8] Neil Spring, Ratul Mahajan, and David Wetherall. Measuring ISP topologies with Rocketfuel. In *Proceedings of the ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication (SIGCOMM)*, pages 133–146, Pittsburgh, PA, August 2002.

[9] Bernard Wong, Ivan Stoyanov, and Emin Gün Sirer. Octant: A comprehensive framework for the geolocalization of internet hosts. In *Proceedings of the ACM/USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, pages 313–326, Cambridge, MA, April 2007.

[10] Ming Zhang, Yaoping Ruan, Vivek Pai, and Jennifer Rexford. How DNS misnaming distorts internet topology mapping. In *Proceedings of the USENIX Annual Technical Conference*, June 2006.

[11] Ying Zhang, Zhuoqing Morley Mao, and Ming Zhang. Detecting traffic differentiation in backbone ISPs with NetPolice. In *Proceedings of the ACM SIGCOMM Internet Measurement Conference (IMC)*, Chicago, November 2009.

7 List of Acronyms

AS	Autonomous System
DNS	Domain Name Service
IAD	Dulles International Airport
ISP	Internet Service Provider
IP	Internet Protocol