ICANN 50

Detecting Distributed DNS Attacks Utilizing Levenshtein String Distances

Nils Clausen, M.Sc.
University Lecturer
n.clausen@ium.edu.na

Detecting Distributed DNS Attacks Utilizing Levenshtein String Distances

Context

Statement of the Problem

Assumptions

The Levenshtein String Distance Measure

Proposed Solution

Sample Result Set

Technical Advice for Implementation

Context

- NA-NIC has turned on the protocol option on their name servers
- the protocol data gets replicated into a relational database (MariaDB)
- table na_log then contains all name server queries with timestamp (down-to-the-second granularity), client ip/port, query name

Statement of the Problem

- NA-NIC has noticed attacks (suspicious queries) on their name servers, possibly caused by bots/viruses and misconfiguration of client networks
- Spikes in query numbers are detected for certain days
- Attacks are not only originating from an easily detectable, uniform range of clients
- Different character permutation techniques seem to be in use by attackers, that makes simple substring comparisons useless for detection

Assumptions

- Suspicious queries:
    - occur only a small number of times per distinct string
    - are systematic and show signs of “somewhat” similarity
    - can be issued from various clients (even at the same time)
    - do not necessarily produce a peak in the number of queries
- Query names, that exactly match registered domains are considered to be legitimate and can therefore be excluded from analysis

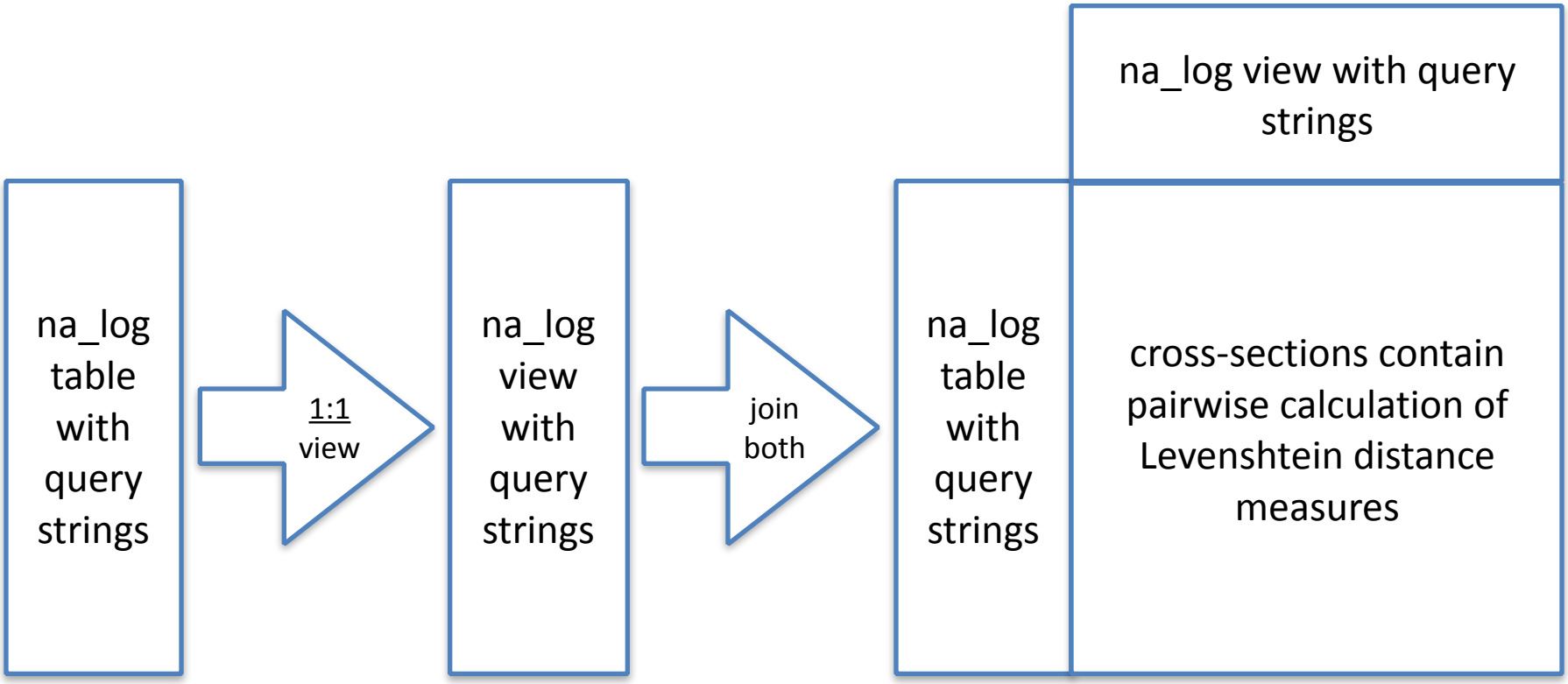
The Levenshtein String Distance Measure

- A string metric for measuring the difference between two sequences, i.e. the minimum number of single-character edits (insertions, deletions or substitutions) to transform one sequence into another
- Levenshtein, Vladimir I. (1966). "Binary codes capable of correcting deletions, insertions, and reversals". Soviet Physics Doklady
- N.B.: used by search engines for suggestions when typing errors are suspected
- Demo: http://odur.let.rug.nl/~kleiweg/lev/

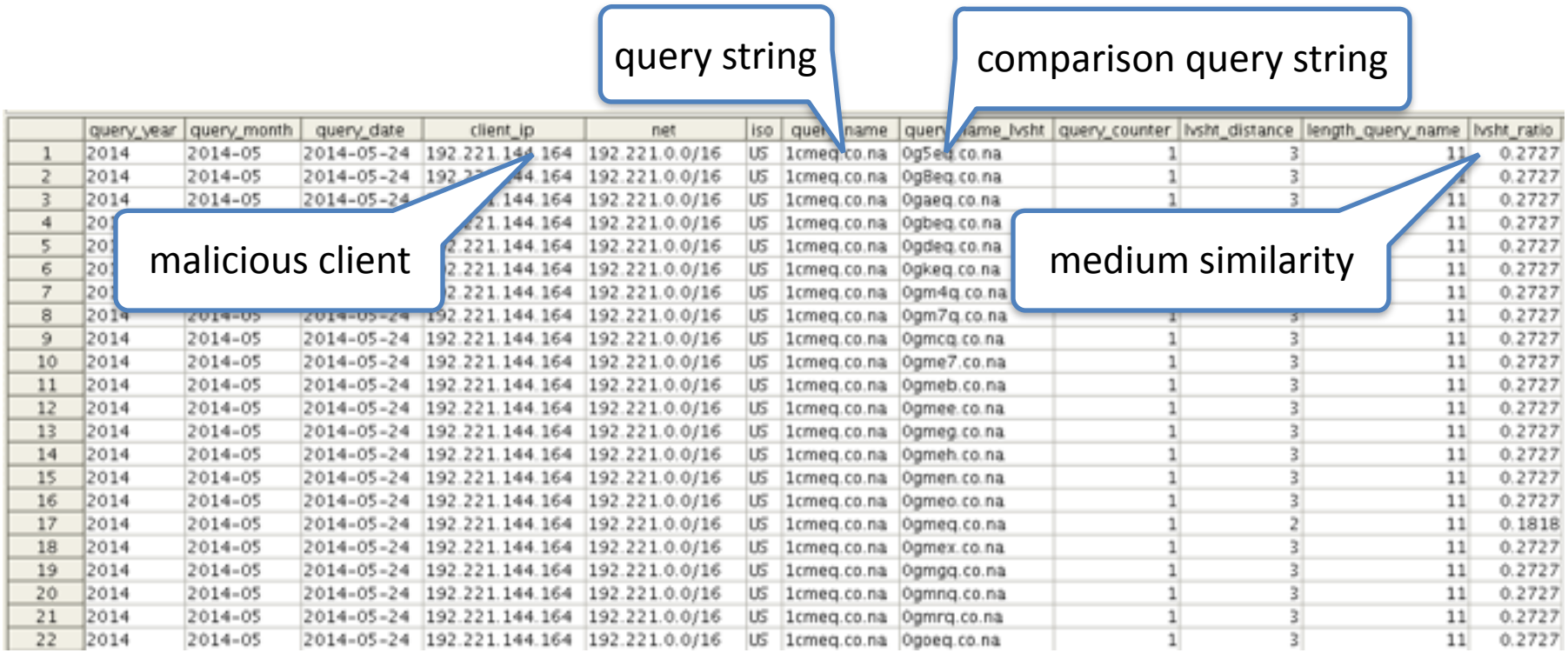
Proposed Solution

- Take na_log as a basis, attributes query_timestamp, client_ip and query_name are of primary interest
- Do pairwise calculation of Levenshtein distances between all query_name combinations with same length
- Limit pairwise calculation to Levenshtein ratios (Levenshtein distance ÷ length of string) > 0 and < 0.3 (to exclude same-string comparisons and only include strings with high- to medium similarity)
- Derive aggregate attributes day, month, year from query_timestamp for further analysis capabilities
- Further calculations can be performed on result set, e.g. correlation metrics for cluster analysis

Illustration of Proposed Solution



Sample Result Set



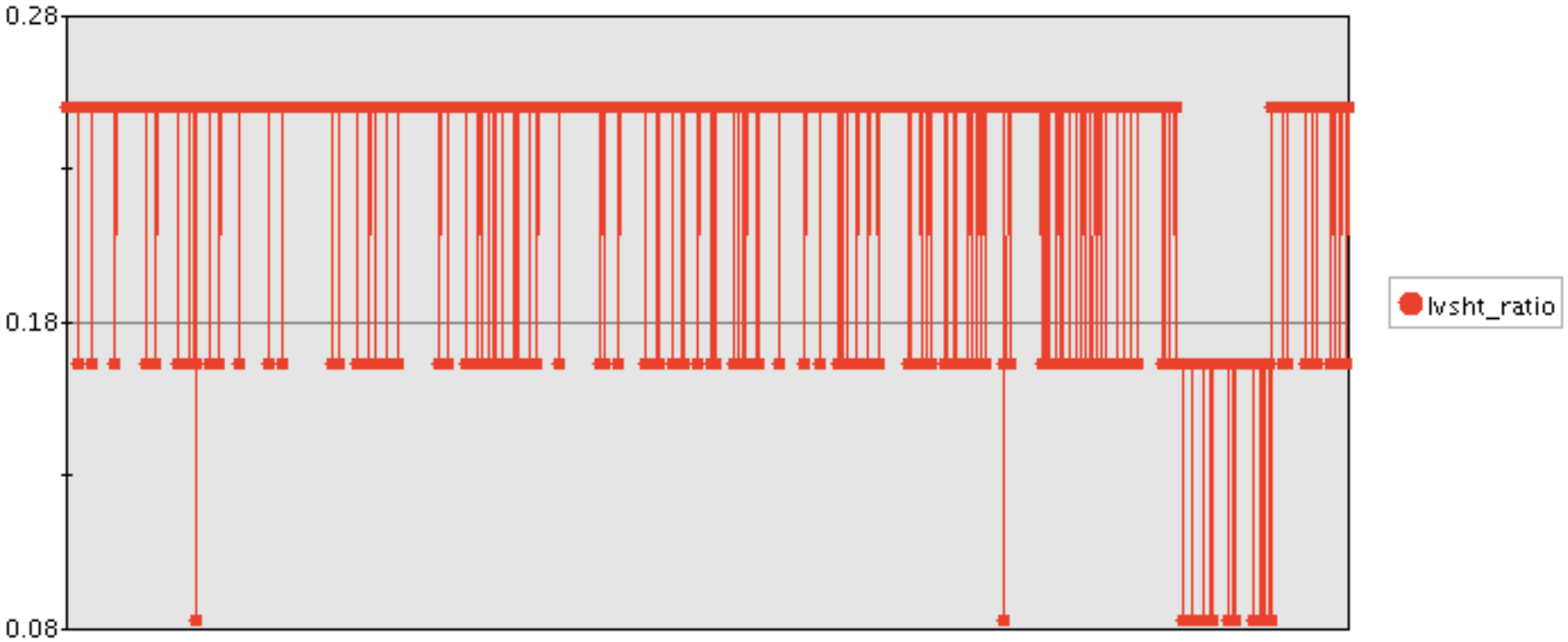
	query_year	query_month	query_date	client_ip	net	iso	query_name	query_name_lvsht	query_counter	lvsht_distance	length_query_name	lvsht_ratio
1	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0g5eq.co.na	1	3	11	0.2727
2	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gBeq.co.na	1	3	11	0.2727
3	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gaeq.co.na	1	3	11	0.2727
4	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gbeq.co.na	1	3	11	0.2727
5	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gdeq.co.na	1	3	11	0.2727
6	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gkeq.co.na	1	3	11	0.2727
7	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gm4q.co.na	1	3	11	0.2727
8	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gm7q.co.na	1	3	11	0.2727
9	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gmcq.co.na	1	3	11	0.2727
10	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gme7.co.na	1	3	11	0.2727
11	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gmeb.co.na	1	3	11	0.2727
12	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gmee.co.na	1	3	11	0.2727
13	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gmeg.co.na	1	3	11	0.2727
14	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gmeh.co.na	1	3	11	0.2727
15	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gmen.co.na	1	3	11	0.2727
16	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gmeo.co.na	1	3	11	0.2727
17	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gmeq.co.na	1	2	11	0.1818
18	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gmex.co.na	1	3	11	0.2727
19	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gmgq.co.na	1	3	11	0.2727
20	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gmnq.co.na	1	3	11	0.2727
21	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0gmrq.co.na	1	3	11	0.2727
22	2014	2014-05	2014-05-24	192.221.144.164	192.221.0.0/16	US	1cmeq.co.na	0goeq.co.na	1	3	11	0.2727

Sample Result Set: Further Analysis

	A	B
		2014-05
	192.221.144.134	183
	192.221.144.164	15
	192.221.144.253	11
	193.226.61.1	81
	193.230.161.3	18
	193.230.161.4	23
	193.230.183.201	111
	193.231.236.17	135
	216.66.80.30	21
	216.66.87.14	43
	8.0.16.101	33
	8.0.16.151	26
	8.0.16.209	93
	8.0.18.13	23
	8.0.18.164	83
	8.0.18.28	25
	8.0.18.58	14

red: high-volume malicious clients, identified by group-by statement on result set

Range of Levenshtein Ratios in Sample Set



Technical Advice for Implementation

- Pairwise comparison implies exponential cardinality of result sets and long running calculation times
- Either limit input to a few hours or days of logs based on database performance, or use in-memory database technology
- Use materialized views or tables to store intermediate result sets for faster access when using non-in-memory databases

Technical Advice for Implementation

- Download levenshtein.c
- Compile as per the file
- Install into MariaDB/MySQL Plugin Directory
- CREATE FUNCTION levenshtein RETURNS INT SONAME 'levenshtein.so';

Thank You.
Questions?