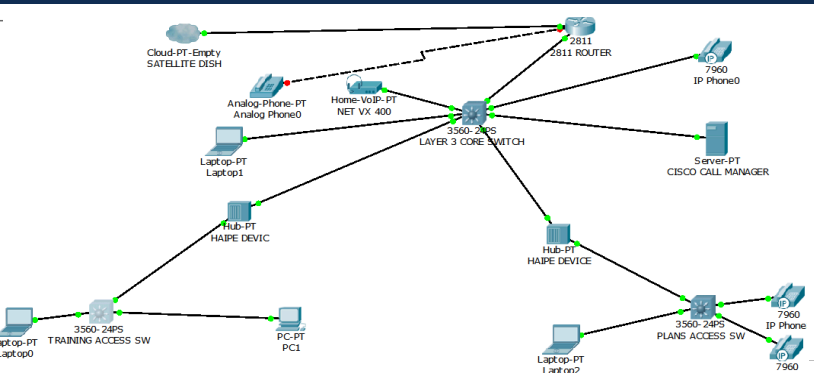
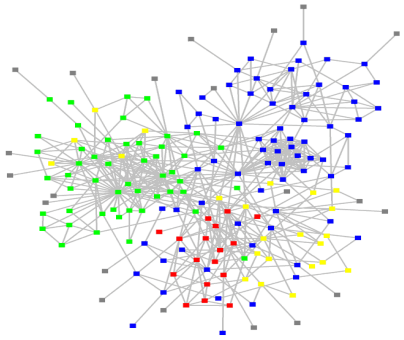
Exceptional service in the national interest

SAND2013-5904C

Sandia National Laboratories





Network Anomaly Detection

Dylan Hutchison

Advisor: Levi Lloyd

Summer 2013

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND NO. 2011-XXXXP

Data Sources

Sandia National Laboratories

Application Level

Low Level

Email Attachments

HTTP Traffic

Known Spam Lists

DNS Lookups

Netflow Data

ARP Requests

DNS Mess

Sandia National Laboratories

```
DNS_RESPONSE^RECURSION_DESIRED^RECURSION_AVAIL^(DNSID)63325^(AN_COUNT)1^(NS_COUNT)2^(AR_COUNT)3^(QREC){tuplebegin}^(QH
OSTNAME)ns2.bdm.microsoftonline.com^(QTYPE)1^(QCLASS)1^{tupleend}^(ANREC){tuplebegin}^(RHOSTNAME)ns2.bdm.microsoftonli
ne.com^(TTL)60^(AIP)157.56.81.41^{tupleend}^(NSREC){tuplebegin}^(RHOSTNAME)ns2.bdm.microsoftonline.com^(TTL)60^(NS)ns1
.bdm.microsoftonline.com^{tupleend}^(NSREC){tuplebegin}^(RHOSTNAME)ns2.bdm.microsoftonline.com^(TTL)60^(NS)ns2.bdm.mic
rosoftonline.com^{tupleend}^(ARREC){tuplebegin}^(RHOSTNAME)ns1.bdm.microsoftonline.com^(TTL)60^(AIP)207.46.15.51^{tupl
eend}^(ARREC){tuplebegin}^(RHOSTNAME)ns1.bdm.microsoftonline.com^(TTL)60^(AAAA)2a01^111^f506^1804^^59^{tupleend}^(ARRE
C){tuplebegin}^(RHOSTNAME)ns2.bdm.microsoftonline.com^(TTL)60^(AAAA)2a01^111^f506^3403^^42^{tupleend}^(DATETIME)2013.0
5.03 08^05^56.235157^(SRCIP)146.216.89.55^(DSTIP)146.216.89.59^(SRCPORT)53^(DSTPORT)38277^(DNSHASH)17431375934665520654

DNS_RESPONSE^AUTHORITATIVE^(DNSID)41671^(NS_COUNT)4^(AR_COUNT)1^(QREC){tuplebegin}^(QHOSTNAME)mx2.gbe0.com^(QTYPE)28^(
QCLASS)1^{tupleend}^(NSREC){tuplebegin}^(RHOSTNAME)gbe0.com^(TTL)300^(SOA)ns1.gbe0.com^(SOA_RESP)dns.gbe0.com^(SOA_EXP
IRE)604800^{tupleend}^(NSREC){tuplebegin}^(RHOSTNAME)gbe0.com^(TTL)300^(RRSIG)
 00 06 08 02 00 00 0e 10  51 8a e7 00 51 78 72 00   ........Q...Qxr.
 19 b0 04 67 62 65 30 03  63 6f 6d 00 70 10 57 8a   ...gbe0.com.p.W.
 6f d8 4c 4f d0 ff cb 95  2c 02 5b 0d 7d 63 f4 78   o.LO....,.(.}c.x
 6c 26 12 ee 50 cc 0a c4  32 b4 68 ac ef 22 6d 4b   l&..P...2.h.."mK
 81 1b 45 5c 5d e5 d7 da  b1 30 63 5f 6e 9b 59 1e   ..E\)....0c_n.Y.
 4d eb 6a 6f 10 81 50 c4  69 4d 49 8a a3 b6 6f d7   M.jo..P.iMI...o.
 00 f4 57 91 19 cd 94 d1  5c 14 80 1d 32 f6 f6 c4   ..W.....\...2...
 0e 7a 8d 5a a2 a4 f5 3a  82 62 b2 4e ba c2 d0 a2   .z.Z...^.b.N....
 e3 ff 67 96 3b 75 e8 95  49 4d ae 5b 26 6d e3 3a   ..g.;u..IM.(&m.^
 78 61 04 9a 35 d1 93 08  50 90 55 66               xa..5...P.Uf
^{tupleend}^(NSREC){tuplebegin}^(RHOSTNAME)p7blq6ts6blhqeclmj8srv6tjgta2itj.gbe0.com^(TTL)300^(NSEC3)
 01 01 00 05 05 27 80 1b  c7 19 14 c9 d7 5d 1b bc   .....'.......)..
 32 eb 1d 39 95 b4 d1 cd  fc dd 9c 3a a1 4b b4 00   2..9.......^.K..
 06 40 00 00 00 00 02                               .@.....
^{tupleend}^(NSREC){tuplebegin}^(RHOSTNAME)p7blq6ts6blhqeclmj8srv6tjgta2itj.gbe0.com^(TTL)300^(RRSIG)
 00 32 08 03 00 00 01 2c  51 8a e7 00 51 78 72 00   .2.....,Q...Qxr.
 19 b0 04 67 62 65 30 03  63 6f 6d 00 ae c2 af ce   ...gbe0.com.....
 55 76 9b 9d 59 ef 1a b5  01 f8 52 1e f8 4b 59 39   Uv..Y.....R..KY9
 2e ac 19 68 89 a1 df c5  16 08 94 1b 40 c3 2d 9e   ...h........@.-.
 8a 1a bc 7a b4 ab bd 74  3b a1 5a 28 e1 b8 3c 27   ...z...t;.Z(..<'
 6b 9b ba b2 73 b0 97 7a  9f dc fb 46 64 84 c8 39   k...s..z...Fd..9
 dc 90 c8 0a 12 a2 97 9f  87 ce c7 60 7b e1 9e e1   ...........`{...
 45 58 56 35 1b 75 42 88  7f ee b2 28 d9 47 81 42   EXV5.uB....(.G.B
```

DNS Mess

Sandia National Laboratories

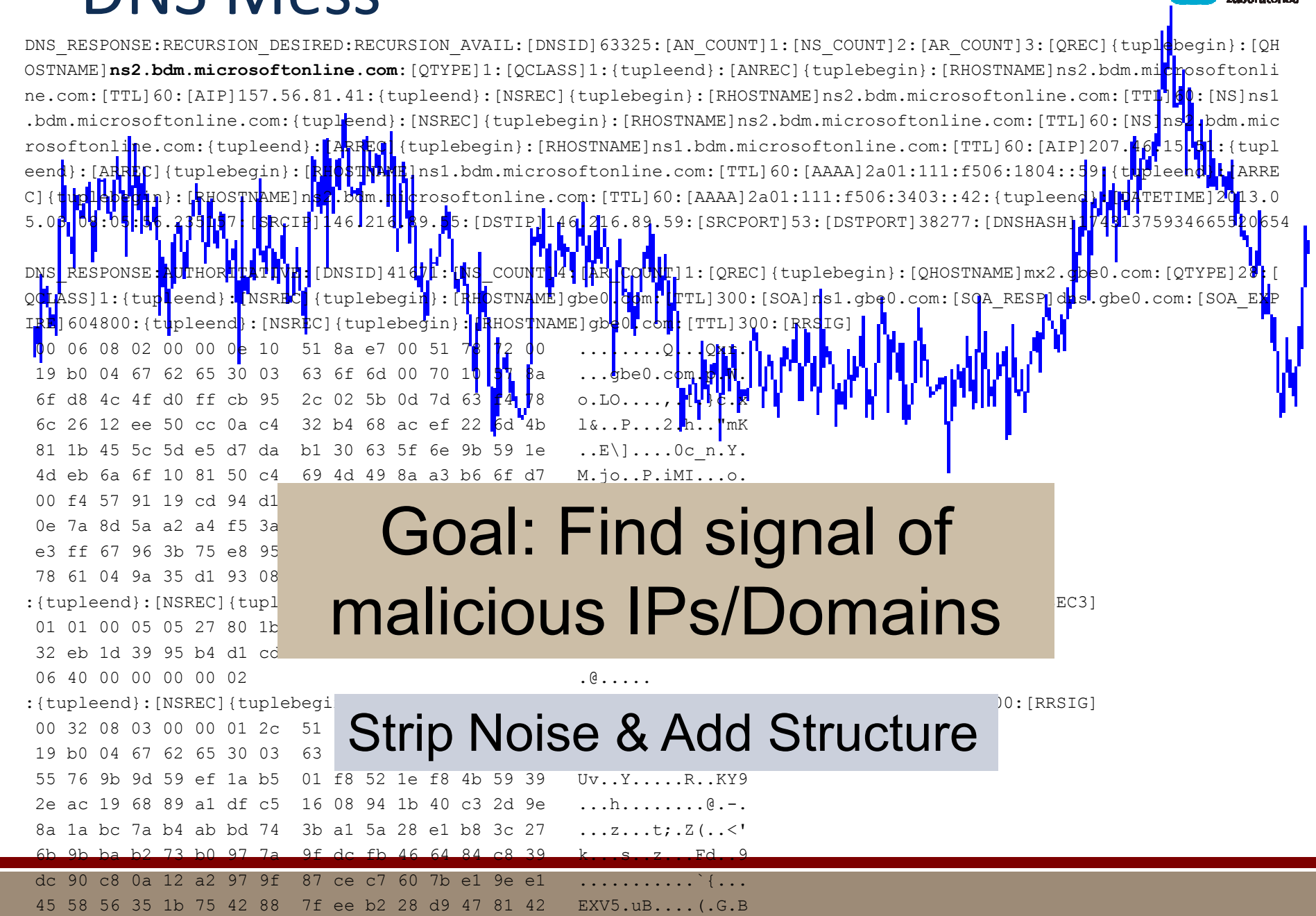
```
DNS_RESPONSE:RECURSION_DESIRED:RECURSION_AVAIL:[DNSID]63325:[AN_COUNT]1:[NS_COUNT]2:[AR_COUNT]3:[QREC]{tuplebegin}:[QH
OSTNAME]ns2.bdm.microsoftonline.com:[QTYPE]1:[QCLASS]1:{tupleend}:[ANREC]{tuplebegin}:[RHOSTNAME]ns2.bdm.microsoftonli
ne.com:[TTL]60:[AIP]157.56.81.41:{tupleend}:[NSREC]{tuplebegin}:[RHOSTNAME]ns2.bdm.microsoftonline.com:[TTL]60:[NS]ns1
.bdm.microsoftonline.com:{tupleend}:[NSREC]{tuplebegin}:[RHOSTNAME]ns2.bdm.microsoftonline.com:[TTL]60:[NS]ns2.bdm.mic
rosoftonline.com:{tupleend}:[ARREC]{tuplebegin}:[RHOSTNAME]ns1.bdm.microsoftonline.com:[TTL]60:[AIP]207.46.15.51:{tupl
eend}:[ARREC]{tuplebegin}:[RHOSTNAME]ns1.bdm.microsoftonline.com:[TTL]60:[AAAA]2a01:111:f506:1804::59:{tupleend}:[ARRE
C]{tuplebegin}:[RHOSTNAME]ns2.bdm.microsoftonline.com:[TTL]60:[AAAA]2a01:111:f506:3403::42:{tupleend}:[DATETIME]2013.0
5.03 08:05:56.235157:[SRCIP]146.216.89.55:[DSTIP]146.216.89.59:[SRCPORT]53:[DSTPORT]38277:[DNSHASH]17431375934665520654

DNS_RESPONSE:AUTHORITATIVE:[DNSID]41671:[NS_COUNT]4:[AR_COUNT]1:[QREC]{tuplebegin}:[QHOSTNAME]mx2.gbe0.com:[QTYPE]28:[
QCLASS]1:{tupleend}:[NSREC]{tuplebegin}:[RHOSTNAME]gbe0.com:[TTL]300:[SOA]ns1.gbe0.com:[SOA_RESP]dns.gbe0.com:[SOA_EXP
IRE]604800:{tupleend}:[NSREC]{tuplebegin}:[RHOSTNAME]gbe0.com:[TTL]300:[RRSIG]
 00 06 08 02 00 00 0e 10  51 8a e7 00 51 78 72 00   ........Q...Qxr.
 19 b0 04 67 62 65 30 03  63 6f 6d 00 70 10 57 8a   ...gbe0.com.p.W.
 6f d8 4c 4f d0 ff cb 95  2c 02 5b 0d 7d 63 f4 78   o.LO....,.[.}c.x
 6c 26 12 ee 50 cc 0a c4  32 b4 68 ac ef 22 6d 4b   l&..P...2.h.."mK
 81 1b 45 5c 5d e5 d7 da  b1 30 63 5f 6e 9b 59 1e   ..E\]....0c_n.Y.
 4d eb 6a 6f 10 81 50 c4  69 4d 49 8a a3 b6 6f d7   M.jo..P.iMI...o.
 00 f4 57 91 19 cd 94 d1  5c 14 80 1d 32 f6 f6 c4   ..W.....\...2...
 0e 7a 8d 5a a2 a4 f5 3a  82 62 b2 4e ba c2 d0 a2   .z.Z...:.b.N....
 e3 ff 67 96 3b 75 e8 95  49 4d ae 5b 26 6d e3 3a   ..g.;u..IM.[&m.:
 78 61 04 9a 35 d1 93 08  50 90 55 66               xa..5...P.Uf
:{tupleend}:[NSREC]{tuplebegin}:[RHOSTNAME]p7blq6ts6blhqeclmj8srv6tjgta2itj.gbe0.com:[TTL]300:[NSEC3]
 01 01 00 05 05 27 80 1b  c7 19 14 c9 d7 5d 1b bc   .....'.......]..
 32 eb 1d 39 95 b4 d1 cd  fc dd 9c 3a a1 4b b4 00   2..9.......:.K..
 06 40 00 00 00 00 02                               .@.....
:{tupleend}:[NSREC]{tuplebegin}:[RHOSTNAME]p7blq6ts6blhqeclmj8srv6tjgta2itj.gbe0.com:[TTL]300:[RRSIG]
 00 32 08 03 00 00 01 2c  51 8a e7 00 51 78 72 00   .2.....,Q...Qxr.
 19 b0 04 67 62 65 30 03  63 6f 6d 00 ae c2 af ce   ...gbe0.com.....
 55 76 9b 9d 59 ef 1a b5  01 f8 52 1e f8 4b 59 39   Uv..Y.....R..KY9
 2e ac 19 68 89 a1 df c5  16 08 94 1b 40 c3 2d 9e   ...h........@.-.
 8a 1a bc 7a b4 ab bd 74  3b a1 5a 28 e1 b8 3c 27   ...z...t;.Z(..<'
 6b 9b ba b2 73 b0 97 7a  9f dc fb 46 64 84 c8 39   k...s..z...Fd..9
 dc 90 c8 0a 12 a2 97 9f  87 ce c7 60 7b e1 9e e1   ...........`{...
 45 58 56 35 1b 75 42 88  7f ee b2 28 d9 47 81 42   EXV5.uB....(.G.B
```

DNSSEC

DNS Mess

Sandia National Laboratories



Goal: Find signal of malicious IPs/Domains

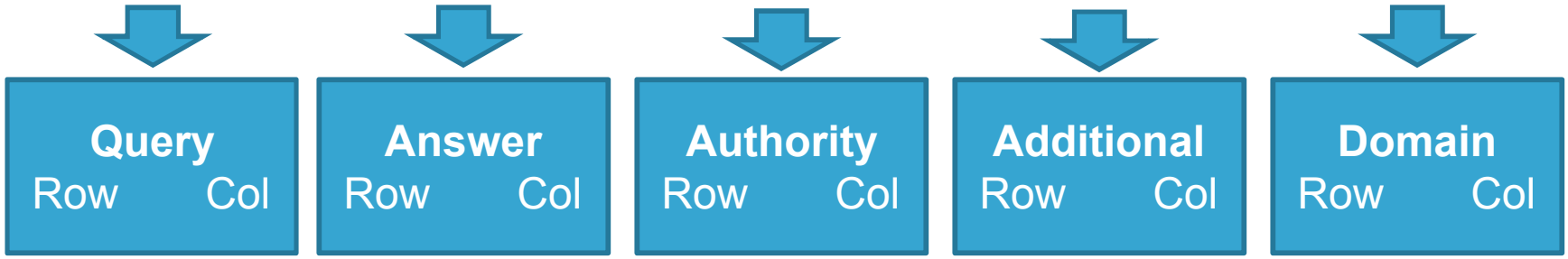
Strip Noise & Add Structure

DNS Data: Parsing

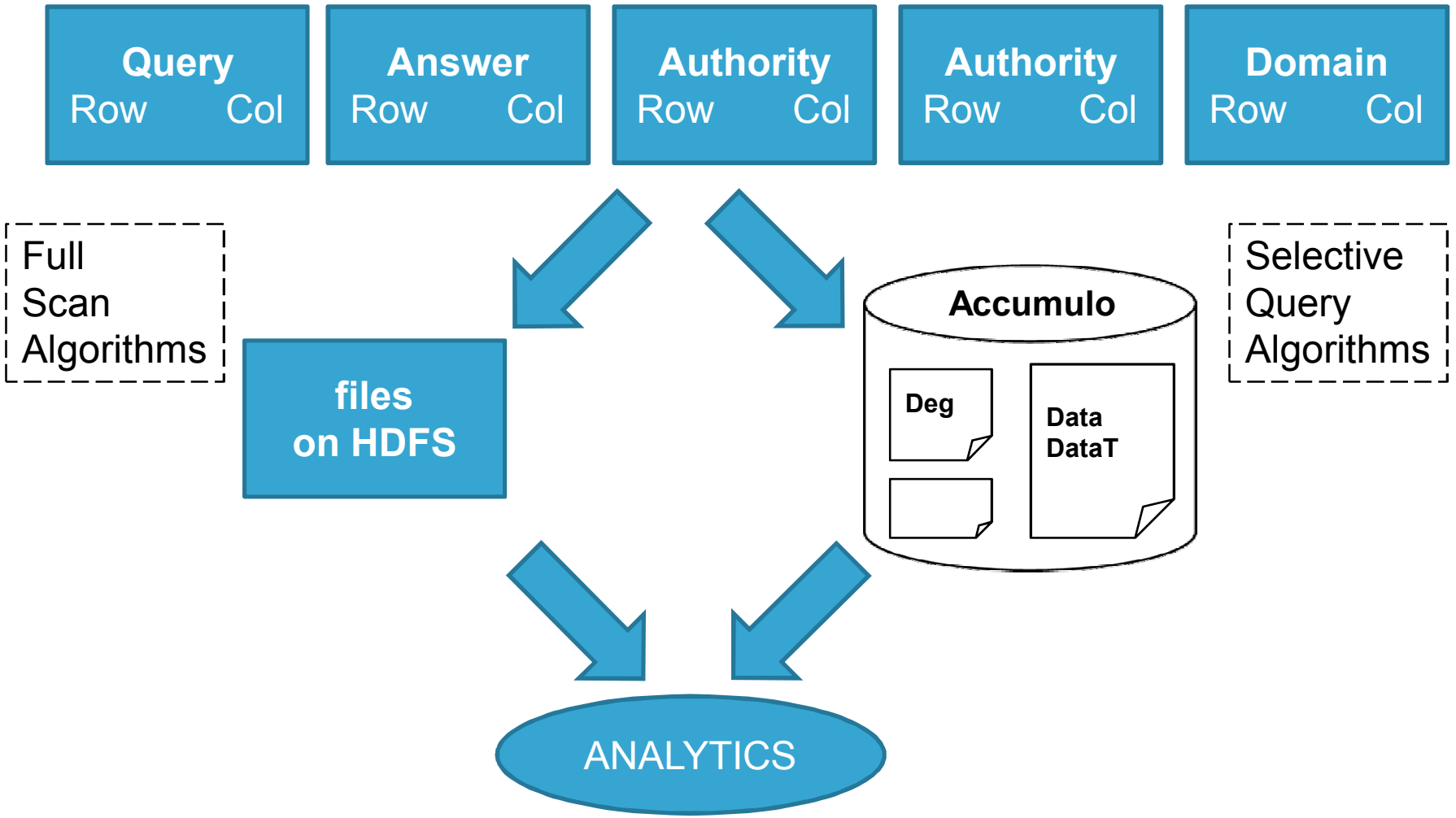
Sandia National Laboratories

```
HEADER:[DNSKEY]15518157687728486961:[DATETIME]2013.05.03
08:05:56.139180:[SRCIP]146.216.89.57:[DSTIP]146.216.89.59
QUESTION:[DNSKEY]15518157687728486961:[QCLASS]1:[QTYPE]1
:[QHOSTNAME]megan.madtech.cx
ANSWER:[DNSKEY]15518157687728486961:[RECKEY]18237466591585589194
:[RHOSTNAME]megan.madtech.cx:[AIP]46.19.34.218:[TTL]2952
AUTHORITY:[DNSKEY]15518157687728486961:[RECKEY]12619112080374549506
:[RHOSTNAME]madtech.cx:[NS]megan.madtech.cx:[TTL]2952
AUTHORITY:[DNSKEY]15518157687728486961:[RECKEY]8405325770053052466
:[RHOSTNAME]madtech.cx:[NS]puck.nether.net:[TTL]2952
ADDITIONAL:[DNSKEY]15518157687728486961:[RECKEY]727239043310050138
:[RHOSTNAME]puck.nether.net:[AIP]204.42.254.5:[TTL]64572
ADDITIONAL:[DNSKEY]15518157687728486961:[RECKEY]10755963813037278145
:[RHOSTNAME]puck.nether.net:[AAAA]2001:418:3f4::5:[TTL]64572
ADDITIONAL:[DNSKEY]15518157687728486961:[RECKEY]2507419571154372886
:[RHOSTNAME]megan.madtech.cx:[AAAA]2a02:2770::21a:4aff:fed1:af3e:[TTL]2952
```

1. Strip noise
2. Structure



DNS Data: Table Ingest



Accumulo Schema

Sandia National Laboratories

Exploded Schema	src_ip|10.0.2.8	src_ip|192.168.1.36	Qhost|.com.google	Qhost|.ly.bit	Qhost|.com.yahoo
lookup_hash|001	1	1	0	0	1
lookup_hash|002	0	1	1	0	0
lookup_hash|003	1	0	0	1	0

- Indexing Across all Variables
- Server-Side Computation
  - Automatic Degree Counts (separate table)

➔ Flexible Analytics

Queries

Sandia National Laboratories

- What are the 4 most frequently returned answers? What were the original requests?
- Find all lookups with TTLs < 10
- Find 10 domains with most unique associated IPs
  - For IPv4, IPv6, both
- Find the top 10 queried domains on May 1, 9:10–9:19

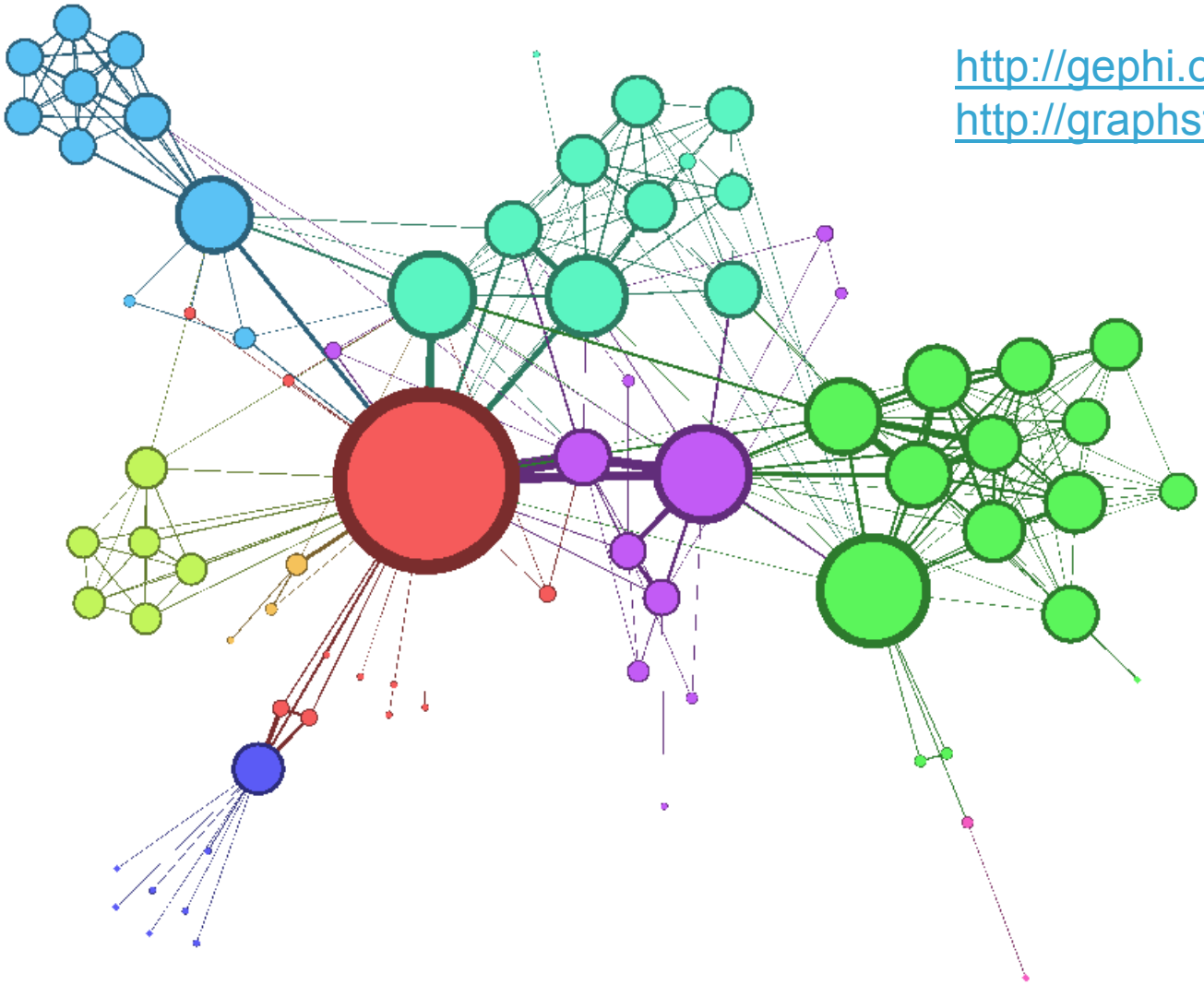
Future Work

Sandia National Laboratories

- Flexibility first, to identify features
  1. Expert Analysis
  2. Machine Learning
- Efficiency next, specializing on key features
- Run on real data!

New angle: Visualization

Sandia National Laboratories



http://gephi.org/
http://graphstream-project.org/