# Touring DNS Open Houses for Trends and Configurations

Andrew J. Kalafut, Craig A. Shue, and Minaxi Gupta

***Abstract*—DNS is a critical component of the Internet. It maps domain names to IP addresses and serves as a distributed database for various other applications, including mail, Web, and spam filtering. This paper examines DNS *zones* in the Internet for diversity, adoption rates of new technologies, and prevalence of configuration issues. To gather data, we sweep 60% of the Internet's domains in June - August 2007 for *zone transfers*. 6.6% of them allow us to transfer their complete information. Surprisingly, this includes a large fraction of the domains deploying DNSSEC. We find that DNS zones vary significantly in size and some span many ASes. Also, while anti-spam technologies appear to be getting deployed, the adoption rates of DNSSEC and IPv6 continue to be low. Finally, we also find that carelessness in handing DNS records can lead to reduced availability of name servers, email, and Web servers. This also undermines anti-spam efforts and the efforts to shut down phishing sites or to contain malware infections.**

***Index Terms*—Domain Name System, Measurement**

## I. Introduction

The Domain Name System (DNS) serves as an Internet-wide distributed database. It maps human-friendly domain names to IP addresses and provides support for applications ranging from simple mail delivery to advanced applications, such as spam filtering, voice over IP (VoIP), and other multi-media services. A typical unit of administration in DNS is a second-level domain name, such as `example.com`. A *zone* file corresponding to the zone stores information about the hosts, services, and sub-domains contained in that zone. While typical DNS queries inquire about a single host or service, some use-cases require complete information contained in a DNS zone. An instance of this occurs when DNS servers for a domain need to synchronize with each other in their view of the zone. The DNS provides a special query for that, called the *zone transfer* query. In this work, we leverage the zone transfer query to capture detailed information about DNS zones in the Internet. During a three month period, we swept 60% of the Internet for zone transfers. In order to increase our data beyond those zones allowing zone transfer, we *walked* the zones of the second-level domains known to deploy DNSSEC [1] (DNS Security Extensions). This is a slow process since it involves making a large number of queries, but its net effect is the same as a zone transfer.

Our work differs significantly from the existing work in the area of understanding DNS zone provisioning and configuration. While the existing work uses a limited portion of the data available at the name servers and focuses primarily on various aspects of name server availability, we instead take a comprehensive look at all the data contained at the name servers. In this study, we focus on analyzing the data from three perspectives: 1) characterizing diversity of zones in the Internet in terms of number of hosts, the domains, autonomous systems (ASes) and BGP prefixes to which they belong, 2) tracking the deployment of new technologies, including DNSSEC, IPv6, and anti-spam technologies, and 3) analyzing zone configuration from the perspective of the availability of various servers, including name servers, Web servers, and mail servers. The key findings of our study are the following:

**Zone transfers:** 6.6% of the second-level domain names in `.com` and `.net` top level domains (TLDs) allowed us to perform a zone transfer of their zones in spite of the well-known fact that the zone transfers are a security risk [2]. Surprisingly, this included a large percentage of DNSSEC-deploying zones, who may be expected to be more careful about security issues.

**Zone diversity:** Zones varied vastly in sizes, with the biggest zone containing over two million hosts when a large fraction contain just a handful. Also, while over half the zones were contained in a single AS, one zone spanned 1,475 different ASes.

**Deployment of new technologies:** DNS-based anti-spam technologies are gaining traction but deployment of DNSSEC and IPv6 continues to be very low. Specifically, 8-16% of the zones deployed DNS-based anti-spam technologies. However, a small fraction of these made mistakes in configuring the relevant records. Fortunately, the email programs at the recipients can be enhanced to account for these mistakes without rendering the deployment efforts ineffective. Only 0.18% of zones in our data deployed IPv6, and only 0.003% used DNSSEC.

**Configuration issues:** Configuration problems were found in many zones. Fortunately, most of these had just one, and in many cases not a serious one. We noticed the possibility of diminished name server availability because although most zones had at least the prescribed two DNS servers, 82% of those observed had both in the same AS. However, the zones deploying DNSSEC had them in the same AS only

Manuscript received August 25, 2008; revised January 29, 2009 and February 26, 2011; accepted March 6, 2011.

Andrew Kalafut (kalafuta@gvsu.edu) is with the School of Computing and Information Systems, Grand Valley State University, Allendale, MI 49401.

Craig Shue (cshue@ornl.gov) is with the Cyberspace Sciences and Information Intelligence Research Group, Oak Ridge National Laboratory, Oak Ridge, TN 37830.

Kalafut and Shue participated in this work while Ph.D. students at Indiana University.

Minaxi Gupta (minaxi@cs.indiana.edu) is with the School of Informatics and Computing, Indiana University, Bloomington, IN 47401

7% of the time. Also, 0.5-11% of zones were likely using the same DNS server for internal and external clients, which is recommended against for security reasons. We also saw misconfigurations that could impact the availability of mail, Web, and other servers, zones exposing more information than likely necessary, as well as zones lacking proper contact information.

The rest of this paper proceeds as follows. Section II provides background on DNS. In Section III, we describe our data collection process, and the issues we ran into in the course of collecting the necessary data. Section IV describes data sanitization and characterizes the zones in both data sets. The analysis of the data contained in zones is presented in Sections V, VI, and VII. We survey related work in Section VIII and conclude the paper in Section IX.

## II. Background

The behavior of the DNS is specified in a series of Internet Engineering Task Force (IETF) Request for Comments (RFC) documents, dating back to the 1980s. While there are many DNS-related RFCs, the key RFCs are RFC 1034 [3] and 1035 [4].

The DNS is organized as a tree, with branches at each level separated by a ".". The entire DNS space is divided into various *zones*. Each zone consists of a connected portion of this tree under the same administrative control. A typical unit of administration in DNS is a second-level domain name, such as `example.com`. A *zone* file corresponding to this second-level domain name stores information about the hosts, services, and sub-domains contained in that zone.

The data within each zone is stored in the form of *resource records* which consists of four basic parts: a *name*, a *class*, a *type*, and *data*. All DNS records relating to the Internet are in the `IN` class. 59 different types of records exist for storing various types of data. A zone is defined by two types of records. The first, `SOA` (Start of Authority), indicates the start of a DNS zone. Each zone should have a `SOA` record. The contents of the `SOA` record are the email of an administrator, the domain name of the primary name server, and various timers. The second, one or more `NS` (Name Server) records, also should exist in each zone. These records indicate the set of name servers for the zone and can also indicate the delegation of sub-zones.

Every DNS zone must have at least one name server which serves the DNS records within that zone. Normally, there is more than one name server for a zone, with one being designated as the *primary name server* and any others being designated as *secondary name servers*. A *zone transfer*, initiated by an `AXFR` query is typically used to transfer the zone data from the primary name server for a zone to the secondary name servers. The primary name server typically loads its data from a flat file known as a *zone file*.

## III. Data Collection Methodology and Issues

We use two data sets in this paper. The first, `zone_transfer`, was obtained by attempting to transfer the zones listed in the `.com` and `.net` TLDs. There were 65,101,733 second-level domains in the `.com` zone file and 9,224,482 under `.net` zone file in June 2007, when we started the data collection [5]. Combined, these 74,326,215 domains represented about 58% of the 128 million zones registered at the time [6]. Even though the zones in this data set are geographically diverse, they lack the perspective from the domains registered under other TLDs, particularly those in various country-code TLDs (ccTLDs). Unfortunately, the ccTLDs do not make their zone files available, making this limitation a fundamental one. For each zone, we had the list of name servers. We looked up the IP addresses corresponding to each of these name servers in order to be able to contact them. We used our own custom software, written using the `Net::DNS` Perl library [7], to zone transfer each of these DNS zones in random order. This process took three months, June-August 2007, in part because zone transfers are connection oriented, unlike regular DNS queries which are connectionless. We attempted a zone transfer from each name server for a zone until we either successfully transferred the zone, or the zone transfer failed for all its name servers. Additionally, if two zone transfers from the same IP address failed, or upon request from the DNS server's administrator, we discontinued making further attempts to transfer any zone from that IP address. Upon connection establishment failure, we retried once. In order to expedite the process, we used five machines, each with one hundred processes issuing zone transfer requests. We succeeded in transferring zones for 4,947,993 (6.6%), indicating that many DNS servers willingly distribute their information to outsiders.

One might argue that the `zone_transfer` data set represents zones that are less security conscious since they allow a zone transfer in the first place. To attempt to compensate for this limitation, we collect a second data set, `dnssec`. This data set is from zones that deploy DNSSEC [1]. DNSSEC adds security to the DNS. These zones may therefore be considered more security conscious, although we note that most of these allowed zone transfer, calling their security practices into some question. DNSSEC provides origin authentication and integrity to DNS data, and authenticated denial of existence. We obtained the `dnssec` data set through walking DNSSEC records. This process is slow but allows retrieval of all the records in a zone, just like a zone transfer does. To build this data set, we began with a list of 862 zones with DNSSEC in production usage from the SecSpider DNSSEC Monitoring Project [8]. We limited this to the second level zones within the `.com` and `.net` TLDs to allow a fair comparison with the zones we transferred data from in the same TLDs. This yielded a total of 124 zones. Surprisingly, we also found 161 zones deploying DNSSEC in our zone transfer data. There was considerable overlap: 96 of the zones listed under SecSpider already existed in our zone transfer data, yielding only 28 new zones. To obtain data from the 28 new zones in the SecSpider data, we used the DNSSEC Walker tool [9]. This tool relies on the presence of `NSEC` (NextSECure) or `NXT` (NeXT) records which should be present in zones deploying DNSSEC. These records provide a way to discover all of the records from within a zone without using zone transfer. Of the 28 zones we attempted to walk, 4 were only partially walkable due to

missing some NSEC or NXT records. The remaining 24 were completely walkable allowing us to get the same information as we would though zone transfer without actually using the zone transfer query. Our final dnssec data set consists of 189 zones: the 161 DNSSEC deploying zones from our zone transfers (which we exclude from the zone_transfer data set) and the 28 discovered through SecSpider. The size of this data set is limited by the low deployment of DNSSEC at the time of this study.

### A. (Non-technical) Data Collection Issues

While zone transfers yield valuable information for research purposes, the technique raises practical, ethical, and legal questions. We encountered various reactions to our data collection efforts from the zone administrators. Many of the early requests we received were concerns that a machine had been compromised or that we were otherwise attacking their systems. As the project progressed, we decided to alter the PTR records (used to map IP addresses to domain names) for each of the scanning machines to indicate that they were involved in DNS research and encouraging the administrators to perform a query for the TXT (TeXT) record on the host name for more details. The TXT record is a free-form record, allowing one to put information in any format. This led the them to a web page explaining the project in detail. This page attracted approximately 300 hits while the experiment was on-going. Over half of the administrators that contacted us were supportive of the work, with a few being being quite enthusiastic. A small number of them requested to have their servers exempted from the scanning, which we promptly honored. One administrator seemed surprised that we would perform such queries without prior permission. Further, even after hearing about the research, one administrator was still livid and stated that our entire prefix had been blocked from his network, with the apparent exception of his mail server.

The issue of zone transfers has reached the legal system. In a civil court ruling which occurred after our data collection, a North Dakota civil court decision declared unauthorized zone transfers in that state illegal [10]. While the circumstances in that case were unique, it is clear that such queries can be viewed as controversial. This further raises the bar on collecting and analyzing the type of data we present in this paper.

## IV. Data Sanitization and Overview

We took several steps to sanitize the data. In this section, we highlight these steps and then present an overview of the resulting data. To keep the discussion simple, we treat the data sets as one in this section.

### A. Data Sanitization

All DNS records in our data have the following format: name IN type data, as explained in Section II. We find issues with all three of the variable fields, described here. Unless otherwise noted, we remove the records mentioned in this section from further analysis.

**Odd record types:** Two of the record types we see are not allocated record types. Specifically, six zones contain records with a type of “65,281” and one zone with “666”. The first is within the range set aside for private use [11]; however, it is unclear what function this record type serves. The second is not even in the range of types allocated for private use.

**Obsolete record types:** Three record types found in our data are obsolete. These are MF, MD, and NXT. Of these, the first two relate to email delivery and are recommended to be substituted by the MX record. All zones containing the MD record contained the MF record as well. There were 178 such zones (0.003%). All of them also contain the recommended MX record, implying that these records are inconsequential. The NXT record was used by older versions of DNSSEC. Only three zones in our data had this record, which is recommended to be substituted by the NSEC record. None of these zones carried the recommended NSEC records, suggesting that they are likely using an old version of DNSSEC and have not upgraded.

**Experimental record types:** We find several email-related experimental resource records in our data as well. These include MB, MR, and MG records, which specify mailbox, mail rename, and mail group. A related, non-experimental but infrequently used email record, MINFO also appears in our data. It is used to send mailing list-related error messages. Each of these records were contained in less than 0.005% of the zones. Incidentally, none of the obsolete, experimental, or odd record types are seen in the dnssec data set.

**Repeated records:** 7099 (0.14%) zones have records that are identical in name, type, and data. These extra copies have little effect on the applications retrieving this data, except in cases when zone administrators fail to consistently update all copies. This could lead to unintended incarnations of records being delivered to the clients.

**Empty name field:** 153 zones contain records with an *empty name field*. These records are not accessible by any DNS query aside from zone transfer, since all other queries require the desired name to be specified.

**Invalid comments:** Lines in DNS zone files are commented by putting a semicolon as the first character. Instead of following this syntax, 4,531 (0.09%) zones contain records that begin with a colon, two slashes, or a hash sign. These are likely failed attempts to comment out old records. Though these records are accessible by anyone specifically looking for them, they have little effect on normal DNS operation.

**Repeated zone name:** In a zone file, domain names not ending with a dot character are considered relative to the zone, so the zone name is added on to them. For example, a record containing www.example.com in the zone file instead of www.example.com. will be replaced by www.example.com.example.com. We find that 6037 (0.12%) zones have it in the name portion of the records and 3217 (0.07%) in the data portion. Making this error inconsistently could break intended relationships between multiple records, causing further errors. Therefore, we leave these records as-is for further analysis.

**Unexpected records:** A zone should only contain records whose name is in the zone, except when they have sub-zones, in which case they are required to have A records for their sub-zones' name servers [12], [13]. We find that 23,947 (0.48%) zones contain unnecessary A records for out of zone name servers. Another 7167 (0.14%) contain other out of zone records. Further, 1857 (0.04%) zones have records which belong in a sub-zone, but these are not the required A records for the sub-zones' name server(s). Failure to keep such records up-to-date can disrupt availability at the clients. Further, the presence of these records, which may not match up-to-date copies stored in the zones where they actually belong, could make it hard to estimate the impact of misconfigurations.

### B. Overview of Collected Data

TABLE I
AGGREGATE STATISTICS

| | |
|---|---|
| Total .com/.net zones | 74,326,215 |
| Name servers by name | 1,611,145 |
| Name servers by IP | 820,547 |
| Zones successfully transferred | 4,947,993 |
| Record types defined | 59 |
| Record types seen in data | 42 |
| Valid record types seen in data | 40 |
| Record types seen in $>10$ zones | 31 |
| Walking of DNSSEC zones | 28 |

Table I presents the aggregate statistics about our combined data sets. We see a total of 42 record types, including the invalid, obsolete, and experimental ones. Some, such as SOA (Start Of Authority), NS (Name Server), A (Address), and CNAME (Canonical NAME) are seen in nearly every zone we examine. Interestingly, the SOA record, the only record type absolutely required for a zone to exist, is the only one that we see in every zone. Even the vital NS is not present in $0.2\%$ of zones, even though it is required by the DNS specification, and despite the fact that we know every one of these zones has at least one name server: the one we used to obtain the zone transfer. The next most popular record type is MX (contains the host name and the priority of an email server). Most other record types are much less widely used, some only appearing in a single zone. Figure 1 depicts the number of zones corresponding to each record type that was seen in 10 zones or more. Clearly, there are large differences in the extent of usage of each of these record types. Although our data only contained zones from the .com and .net TLDs, we examined the LOC (LOCation) records for the 1,306 zones which contained them, and found them to be well distributed geographically.

## V. ZONE DIVERSITY

We start by examining the diversity of zones contained across our two data sets. We consider two aspects of zone diversity: their sizes and their span across ASes and BGP prefixes.

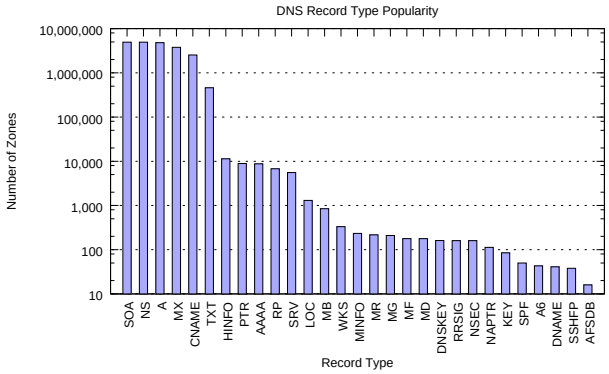



Fig. 1. Number of DNS zones containing popular record types (log scale)

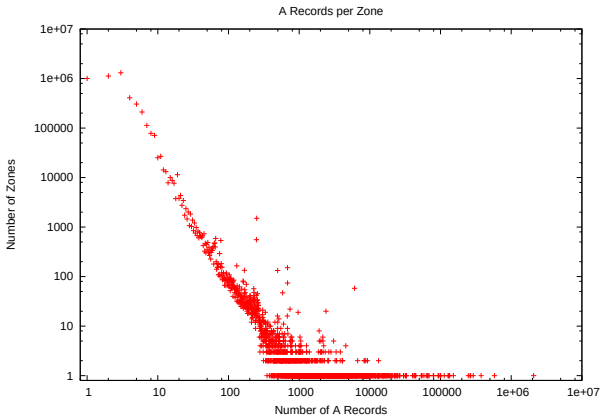



Fig. 2. Number of A records per zone in the combined data set (log-log scale)

### A. Zone Sizes

One approach to looking at zone sizes is to look at the total number of records contained in various zones. However, this approach is dependent on what record types a zone chooses to use. Some records, such as CNAME do not add any new hosts but provide extra information about an existing record. Thus, we count the A records in order to estimate the size of a zone. Since all hosts must have an A record, the number of A records in a zone should roughly correspond to the number of hosts in the zone intended to be accessible though DNS. We ignore the AAAA (IPv6 address) records in counting hosts since very few zones use IPv6 and even when they do, they usually have IPv4 records for the same hosts.

Figure 2 shows the number of A records per zone. As seen in the figure, a majority of zones are small, containing only one A record. Some have more, but it is surprising how much more. The largest has 2,073,715 A records. There are additionally 14 others with over 100,000 A records, although no others with over 1,000,000. The largest zone we see has many A records in part because they have an A record for each address in the 10.32.0.0-10.63.255.255 private IP address space in addition to enumerating every address in another public prefix. Most of rest of the zones with a large number of A records follow either this pattern of an A record for every address in a prefix, or they have a large number of domain names all pointing to the same IP address.

The pattern we see in A records is also seen in other record types, including ones which we do not expect, such as MX. While most domain names with MX records only have a few, we see one domain name with 1,844 MX records pointing to different mail servers. It is unknown why a domain would use so many mail servers as this – even large email providers, such as hotmail.com or yahoo.com each have less than 15 mail servers according to their DNS entries.

### B. Zone Span

We measure zone span by examining the A records from each zone and finding the AS and BGP prefix to which the address belongs. To perform the classification, we use a BGP RIB from the Route Views Project [14] from the same duration as our zone transfers. We use this to determine the number of unique ASes and prefixes the zone entries span. In Table II, we show the breadth of the zones by the AS they belong to. 2.8% of zones are not associated with any AS, meaning all their machines are in private address space, the zone is only used internally but accessible externally. A majority of zones, 56.3%, have A records contained in a single AS. *94% of zones are contained in 2 or fewer ASes.* Only a very small number of zones span more than 4 ASes. A small number of zones were exceptional, however. Specifically, one zone spanned 1,475 ASes, and another 40 spanned 100 or more ASes. The ones with the largest span are dynamic DNS providers. This shows that the zones cover both ends of the spectrum: from tightly co-located networks to highly distributed collections of machines. When analyzing zones at the BGP prefix granularity, we found similar trends. We omit these results for brevity.

TABLE II
NUMBER OF ASES PER ZONE.

| Number ASes Per Zone | Number of Zones | Percent of Zones | Cumulative Percent |
|---|---|---|---|
| 0 | 137,358 | 2.78% | 2.78% |
| 1 | 2,786,918 | 56.32% | 59.10% |
| 2 | 1,881,611 | 38.03% | 97.13% |
| 3 | 114,594 | 2.32% | 99.44% |
| $\geq 4$ | 17,198 | 0.35% | 99.98% |

## VI. DEPLOYMENT OF NEW TECHNOLOGIES

Making infrastructure changes in the Internet is often an uphill battle today. The zone transfer data can provide insights into which new technologies are getting adopted. In this section, we investigate the deployment of DNSSEC and IPv6, which have existed for more than a decade, and a few newer technologies, including SSH fingerprints, spam prevention technologies, and service discovery, which are already starting to be deployed. We do so by looking at the relevant DNS resource records for each of these technologies. It is noteworthy that inferring the adoption of a few of the technologies we describe subsequently, such as service discovery and SSH fingerprints, would be hard to study without the zone transfer data.

### A. DNSSEC Deployment

DNSSEC [1] (DNS Security Extensions) is a set of extensions to the DNS which provide origin authentication and integrity to DNS data, and authenticated denial of existence. The deployment of DNSSEC has been studied previously [8], [15]. However, the previous works relied only on voluntary submission of data from zones deploying DNSSEC. The zone transfers allow us to also learn about zones that may not have reported their deployments. Indeed, we find 65 such zones. This is significant given that the SecSpider project only reported 124 DNSSEC-deploying zones in .com and .net TLDs.

The DNSSEC protocols use four record types, DNSKEY, RRSIG, NSEC, and DS. Any zone deploying DNSSEC must have the DNSKEY record, as it contains the public key used to verify signatures used in DNSSEC. *We only see the DNSKEY record in 161 zones from our zone transfers, which is a mere 0.003% of the zones that allowed us to do a zone transfer*. This corroborates the previous findings that the adoption rate of DNSSEC is extremely low.

Out of the rest of the DNSSEC record types, RRSIG is most important, as it provides signed record types. Without this record type, a zone cannot claim to be deploying DNSSEC. All but one of the 161 zones provide at least one RRSIG record. Expectedly, the same zones that have RRSIG records also have NSEC records. This record allows a traversal of records and is used for authenticated denial of existence. The last record, DS, is used by the zone to authenticate the DNSKEY records of its sub-zones. Only two zones contained this record. Three additional zones contained sub-zones, but did not have DS records. For these three, the zone cannot provide authentication of the sub-zones [16].

### B. IPv6 Deployment

There are various ways in which the adoption of IPv6 can be inferred. One such method is to look at IPv6 address allocations and announcements in the routing protocols. Work in [17] looked at routing announcements and found at most 807 IPv6 prefixes observed at a single location at the start of 2007. This method shows which networks are capable of IPv6, but not who is actually offering service though it. The zone transfers offer a different perspective. They allow us to see how often publicly-accessible servers are available though IPv6. Just like A records provide host name to address mappings for IPv4 addresses, AAAA records provide host name to address mappings for IPv6 [18]. *8,714 zones (0.18%) in our zone_transfer data and 23 (12.2%) in the dnssec data are deploying AAAA records.* Clearly, IPv6 has a long way to adoption.

Examining the AAAA records in detail, we find that zones deploying IPv6 are doing so minimally. A majority of zones, 80.8%, have AAAA records for a subset of the names contained in A records when one would expect that if a zone wanted to make all its hosts accessible by both IPv4 and IPv6 clients, it will have an AAAA record for each A record. In fact, an overwhelming number of the zones with fewer AAAA records than A records only have one AAAA record. Of the rest, a large

majority have disjoint AAAA and A records. These zones are most likely deploying IPv6 only for certain services they know are only going to be accessed by clients from other zones that deploy IPv6. On the other extreme, a few zones, 17, only have AAAA records, but no A records at all. Clearly, no IPv4 client can access them.

### C. Secure Shell (SSH) Fingerprints

The SSH protocol provides secure log-in and other secure network services over an insecure network. The security of the connection relies on the server authenticating itself to the client as well as the client authenticating itself to the server. When a SSH client connects to a server whose public key is not already known to the client, the server presents a fingerprint of the key for verification. If the client accepts the fingerprint (and hence the key), the key is saved locally and used for verification for all subsequent connections. Today, most users blindly accept the presented key. However, the SSHFP record attempts to provide a solution to this problem by providing the fingerprint of server public keys trough DNS [19]. An SSH client can query the DNS for this record and verify the fingerprint before accepting server's public key. *Only 29 zones in our zone_transfer data and 12 (6.3%) in our dnssec data support SSHFP, implying that the adoption of SSH fingerprints is low as of now.*

### D. Anti-spam technologies

Spam is undoubtedly one of the biggest security issues facing the Internet today. To avoid accepting spam, technologies that verify sender identity before accepting email have been proposed. Prominent examples of email verification systems are DomainKeys [20], [21], SenderID [22], and Sender Policy Framework (SPF) [23]. SPF and SenderID help verify that the machine that sent an email was authorized to do so. DomainKeys is a public/private key authentication system which verifies that a message indeed came from the domain it claims and that it has not been modified.

The anti-spam technologies rely on the DNS infrastructure in one of two ways. First, they all have a specially-formatted TXT record. (The TXT record could be used for a variety of other purposes as well.) Second, SPF has a special record type defined for itself, SPF, which was introduced later. We find that SPF is the most popular anti-technology in our DNS zones. 409,214 zones (8.3%) in zone_transfer data and 31 zones (16%) in the dnssec data set used the SPF technology through TXT records. Only 50 zones in the zone_transfer data use the SPF record while none use it in the dnssec data set. Much smaller percentages of zones deployed DomainKeys or SenderID in either of the data sets. *This indicates that a significant fraction of zones in the Internet employ DNS-based anti-spam technologies, with those deploying DNSSEC doing so even more.*

### E. Service Discovery Deployment

There are several different service discovery mechanisms deployed in the DNS. Some services have their own record types, such as MX to find mail servers and AFSDB to locate AFS database servers. However, other more general DNS mechanisms can locate a variety of services. WKS (Well Known Services) and SRV (SeRVice) records both support finding services in different ways. The SRV record specifies both the supported protocol and the port it is running on among other things. We find a total of 89,010 SRV records from 5,548 zones in the zone_transfer data and 9 in the dnssec data. Light-weight Directory Access Protocol (LDAP) alone accounts for 39% of the SRV records. The next one was Kerberos authentication system (combining TCP and UDP), representing about 24% of records. None of the remaining protocols had 1,000 or more entries. The top 5 services found advertised with this record are shown in Table III.

TABLE III
TOP SERVICES IN SRV RECORDS.

| Protocol | Transport | Entries |
|---|---|---|
| ldap | tcp | 35,150 |
| kerberos | tcp | 16,903 |
| gc | tcp | 6,451 |
| kerberos | udp | 4,062 |
| kpasswd | udp | 3,969 |

The WKS record is also used to indicate service availability, but is far less popular than the SRV record. WKS records were present in 331 zone_transfer zones and no dnssec zones. We obtained 1,717 WKS records indicating the availability of 2,751 services. Of these services, FTP was the most commonly advertised at about 19%.

### F. Deployment of Multimedia Services

The NAPTR record is used for URI re-writing, but provides evidence of multimedia service usage. The Session Initiation Protocol (SIP) protocol uses the NAPTR records for providing locator services for Voice over Internet Protocol (VoIP) and other multimedia [24]. Out of the 111 zones in the zone_transfer data containing NAPTR records, 98 are using the NAPTR records to support the SIP protocol. Three zones in the dnssec data are using NAPTR, two of these for SIP.

## VII. Analysis of Zone Configurations

We now look at configuration problems in DNS zone contents. When looking for configuration problems, we only look within each zone independent of others. Since not all zones allow zone transfers, and our transfers were done over a period of three months, we can not accurately identify configuration problems involving the interaction between multiple zones. Because of this, if a record points outside its zone, we assume it to be correct. This means the numbers reported here for many of the misconfigurations are lower bounds, the actual extent of misconfiguration may be higher. We find that while a large fraction of zones have at least one type of misconfiguration, it is uncommon for a zone to have multiple problems simultaneously. Few individual problems occur in large percentage of zones.

### A. Invalid Hosts

The NS records in a zone indicate the name servers for that zone and for its sub-zones. Problems in these records can slow down DNS queries for the zone or even make the sub-zones inaccessible. We find that many zones have NS records that point to host names which are not externally accessible. In our zone_transfer data set, 35,618 zones (0.72%) have NS records with host names consisting of a single label (a host name with no dots in the name). These cannot be a host within any domain because a valid host name must have at least two dots in it. Further, 3,437 zones (0.07%) have NS records indicating name servers with host names in the .local TLD, which is not a valid TLD. Neither of these errors occur in any dnssec zone. We also see problems in the hosts pointed to by the NS records. In 24,457 zones (0.5%) in the zone_transfer data and one zone in the dnssec data, there are NS records pointing to hosts for which no A or CNAME records exist.

We see similar problems in MX records, which are used to indicate the email server for a domain, and in CNAME records, which are used to provide an alias for a host name. In 4,452 zones (0.09%) in the zone_transfer data, there were MX records pointing to a host name which consists of a single label, and in 17 zones in this data, MX records point to mail servers in the .local TLD. The net result of these errors is the unavailability of mail for the domain name of the record if these are the only MX records for a domain, or delays in mail delivery if there are others. As was the case for these problems in NS records, neither of these errors occur in the dnssec data. Many zones with valid MX records have issues with the hosts those records pointed to. In the zone_transfer data, we also found that 18,376 zones (0.37%) had MX records pointing to host names with no A or CNAME records. This issue was seen in 2 of the dnssec zones as well. Looking at CNAME records, in 3,109 (0.06%) zones in the zone_transfer data, the CNAME records point to a host name that is empty, an IP address, a URL instead of a name, in .local TLD, or has a single label instead of at least two. None of these errors occur in CNAME records in the dnssec data.

In CNAME records we also see a few other problems. First, we see chains of CNAME records with one CNAME pointing to another in 28,082 (0.57%) zone_transfer zones and 5 dnssec zones. This has the effect of slowing down DNS resolutions involving these records. In fact, some of these chains have loops: *9970 (0.2%) zones in the zone_transfer data and 1 zone in the dnssec data have loops. These will cause the CNAME to be unresolvable, leading to unavailability.* Further, we also find that 11,414 (0.23%) zones in the zone_transfer data have CNAME records with the same name as another record. This could create ambiguity in the resolution process. This problem was not present in any of the dnssec zones.

### B. Diminished name server redundancy

The NS records also shed light on the name server redundancy provisioned by the zone. Every zone is required to have at least two name servers [3] and recommended to have at least three [25]. This ensures availability of records when attacks or outages occur. 1,665 zones (0.03%) in the zone_transfer data list no name servers at all even though they are required to. Note, however, that this does not make them inaccessible. Clearly, they are accessible since we transferred their zone. Instead, it implies that their NS server records existed in their parent zone, but not in the zone itself, as they are also required to. This problem does not occur in the dnssec data. Further, we find that 11.9% of zones in zone_transfer data list less than the required two name servers. 66% of zones list three and 22.1% list even more. The dnssec zones are provisioned much better with only 3% of the zones with less than the required two name servers.

By separating name servers, both physically and in the network topology, zones can ensure that redundancy provides greater resiliency [25]. We examine name server redundancy at several granularities: according to the BGP prefix advertisements, by autonomous system (AS) they belong to, and across second-level domain names (the final two components of a domain name). Table IV shows that 82% of the name servers in the zone_transfer data set are within the same AS, 61% within the same BGP prefix, and 91% within the same second-level domain. *This implies that the name servers are not physically or topologically distributed for many zones, which may make them susceptible to single points of failure.* Correspondingly, 7% of dnssec zones are in the same AS, 5% in the same prefix, and 12% in the same second-level domain. Clearly, the dnssec zones pay attention to the quality of redundancy in their name servers.

TABLE IV
NUMBER OF ASES, BGP PREFIXES, AND SECOND-LEVEL DOMAINS NAME SERVERS OF THE ZONES CONTAINED IN THE TWO DATA SETS BELONG TO

| # | Percent of Zones | | | | | |
|---|---|---|---|---|---|---|
| | zone_transfer | | | dnssec | | |
| | AS | Prefix | Domain | AS | Prefix | Domain |
| 1 | 82.3% | 61.0% | 90.7% | 6.9% | 4.8% | 12.2% |
| 2 | 15.6% | 22.3% | 8.4% | 87.3% | 33.3% | 84.1% |
| 3 | 1.9% | 3.0% | 0.5% | 3.2% | 58.2% | 2.7% |
| 4 | 0.2% | 13.6% | 0.2% | 2.1% | 3.2% | 0.5% |
| 5 | 0.04% | 0.06% | 0.00% | 0.00% | 0.00% | 0.00% |

### C. Information Leakage

The WKS records help in service discovery. Given a host name, they can find all the services running on that host. This allows an attacker to gain knowledge about all the vulnerable services on a given machine. The vulnerability presented by the WKS records becomes more of a security risk in the presence of HINFO (Host Information) records, which map a host name to its machine type and operating system (OS). *11,379 (0.23%) of the zones in the zone_transfer data set and 2 of the zones in dnssec data set had HINFO records. Of these, 16% give away machine and OS information for the domain's Web server and another 13% including one of the dnssec zones for the email server.* Further, 247 have HINFO records for a host they also have a WKS record for.

While these records can be potentially security risks, it is difficult to aggregate them for analysis due to the fact that

DNS operators are not following any standard convention in the records. Though there are standard values meant to be used for hardware type and OS in the HINFO records [26], [27], 52% and 92% zones are using OS names and hardware types not on the list for at least some of their HINFO records. For example, simply classifying a machine as a "Linux PC" requires automated interpretation that "Linux," "Fedora2," "Slackware 7.0," and "Debian Sarge" are all labels that refer to Linux and that "i386," "Pentium IV-2.4," "IBM-PC," and "P4" are all labels synonymous with a PC. The downside to this is any protocol that may take advantage of HINFO records will likely not be able to use non-standard values. The upside however, it that it would be harder for a malicious application to do this as well.

Not all information leaks come from records designed to expose information about hosts. In 0.5% of zone_transfer zones and 22 (11.6%) dnssec zones, we find A records pointing to private IP addresses. Private IP addresses are only usable on internal networks. Since these records cannot be used by external hosts, their presence in a zone may be an indication that the zone is running the same DNS server for internal and external clients, and not separating them as is recommended. This has the unfortunate consequence of exposing the internal DNS server to attacks when separating the two would normally make it hard for an adversary to even know the whereabouts of the internal DNS server (NS records are for external DNS servers only). Even if the two servers are not combined, the private IP addresses in these records are still a potential problem, as they may be exposing information about which hosts exist on an internal network to external clients with no need for this information.

DNS TXT records are unique in that they allow arbitrary text in their data field. The contents of some TXT records are intended to be interpreted by humans while those of others are specifically designed for machine interpretation. While a small number of TXT records are used for anti-spam systems and DNS-based service discovery, as we discussed in Sections VI-D and VI-E, some of the rest contain sensitive information, such as addresses, telephone numbers, the date a zone was last updated, or which DNS server is the primary DNS and which is secondary. While we saw all of these uses, it is difficult to quantify how often TXT records are used for each due to the free-form nature of the data. Another popular use of the TXT records is to advertise hosting services. Finally, in one case, we found poetry, written as a set of seven TXT records.

### D. Implications on Caching

An SOA (Start of Authority) record indicates the start of a DNS zone. Each zone is required to have a SOA record. Among other things, the SOA records contain the values of the four timers which are important in DNS zone operations. These are the *refresh*, *retry*, and *expire* intervals, and the *minimum TTL*. The refresh, retry, and expire intervals all control the behavior of secondary DNS servers with regards to updates. The refresh interval indicates the amount of time (in seconds) a secondary DNS server should wait before checking

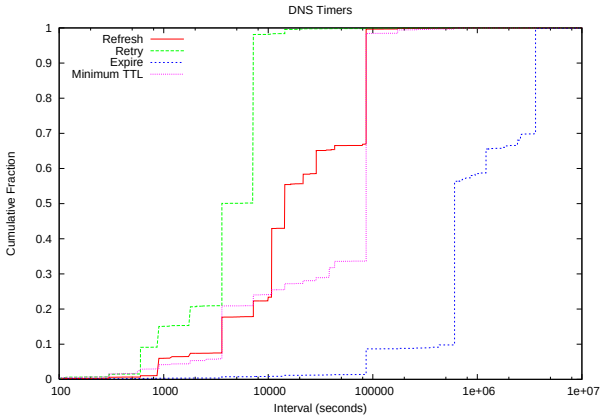



Fig. 3. CDF of refresh, retry, expire, and minimum TTL timers contained in SOA records in the zone_transfer data set (log scale).

to see if its copy of the DNS zone is current. The retry interval indicates how quickly it should retry this operation if it is unsuccessful at the end of the refresh interval. The expire interval indicates the amount of time that can elapse without successfully refreshing the zone before a secondary name server can no longer give authoritative answers to DNS queries for the zone. The minimum TTL is the default duration for which records from this zone can be cached by DNS resolvers.

We find that 14,003 (0.28%) of the zones in the zone_transfer data set and 2 in the dnssec data set have their expire timers set to values less than the refresh timers. This implies that there will be a period where the DNS records cached at the secondary name server will be invalid before they are refreshed. During such a period, the availability of all the secondary servers will be reduced. Further, while the common values used for refresh and retry timers are mostly within the range of those recommended [13], the common values for the expire timer are 7 days and 41.6 hours. Both of these fall outside the recommended interval, which is 2-4 weeks.

Figure 3 shows the CDF of refresh, retry, expire timers, and minimum TTLs seen in our zone_transfer data. One key observation from this data is that some values for these timers are chosen by a large percentage of zones. The common values for the refresh timer are 1 hour, 3 hours, and 1 day when the recommendation is for them to be between 20 minutes to 12 hours. The common values for the minimum TTL timer are 1 hour and 1 day when the recommendation is for 1-5 days. Some zones choose very small or very large values for these timers. While very small values put extra burden on the secondary DNS servers in keeping their view of the zone updated and also on DNS resolvers around the world, very large values can hurt zone availability when records in such a zone change.

### E. Incomplete Contact Information

It is increasingly important that zone administrators be reachable. One example of such importance is phishing, where the process of shutting down phishing sites hosted at compromised servers belonging to reputable domains can benefit from

being able to easily reach the domain administrators. Similarly, isolating members of bot armies or infected machines spreading malware can benefit significantly from the ability to contact their administrators. There are two places in the DNS records where such information is available. The first is the SOA record, which all zones are required to have. We find that all SOA records contain email addresses but 29,946 (0.61%) of the zone_transfer zones have it in an incorrect format: they forget to replace the "@" in the email address by a "." as required. Fortunately, this mistake is easy to account for.

The second place where the information about administrators can be present is the RP (Responsible Person) record. This record contains the email address of the zone administrator and a pointer to a TXT record containing additional information. The email address in RP records should be formatted as in the SOA records. Unfortunately, a very small fraction of zones have this record: Only one dnssec zone and 6770 (0.14%) of the zone_transfer zones have it. Further, 2.6% of the RP records either contain no information or contain a single label that could not be an email address. Another 71.6%, including one from the dnssec data set, just contain the email address and point to an unusable TXT record or a non-existing one. This implies that 3/4th of the RP records at best contain as much contact information as the SOA record.

## VIII. Related Work

Wanrooij *et al.* [28], characterized DNS misconfigurations from a sample of the .NL TLD. They did so by performing DNS ANY queries on 10,000 randomly selected zones mentioned in the .NL zone file. Their study had limited view of DNS provisioning because the ANY query, as they used, provides only a small subset of the records in a zone. Our analysis considers extensive information about orders of magnitude more domains. The richness of DNS records contained in our data sets allowed us to gain a deeper understanding of availability of various kinds of services, and also security implications.

Pappas *et al.* [29] examined the impact of three specific DNS configuration errors: lame delegation (the name server(s) present at the zone differ from those present at the parent zone), diminished server redundancy (less than adequate number of name servers are available or the available servers are not topologically dispersed, implying that they may become unavailable under attack or outage conditions), and cyclic dependency (name servers point to each other, forming a loop). While their work focused on name server availability, we focus more on the availability of other servers, including mail server, Web server, etc. We do consider aspects of name server availability. However, our results are not directly comparable to theirs due to difference in methodology.

The Measurement Factory [30] performed zone transfers on a small fraction of the .com and .net zones. They randomly sampled about 3.22% of .com and .net zones and attempted to transfer them. Though they had data similar to us, they utilized it in ways that differ significantly from us. While we focus on information contained in zone records, they focused on the versions of DNS software in use (to infer possibility of cache poisoning), lame delegation, diminished server redundancy, and possibility of recursion (to infer potential misuse of such name servers by escaping detection). Surprisingly, they find that over 30% of the name servers allow a zone transfer. We find this percentage to be much lower – we were only able to transfer 6.6% of zones out of all the ones we attempted. Another area where we did similar measurements is the adoption of new technologies. We find that the adoption rates of various technologies, including SPF, DNSSEC, and IPv6 are lower than what Measurement Factory reported. Since we do have access to their data, we conjecture that the differences in the numbers arise out of sampling. In our previous work [31], we presented a limited perspective on DNS configuration issues in a short paper. This paper extends that work by examining diversity of DNS zones, by investigating the deployment of new technologies, and by presenting a more thorough investigation of configuration issues.

A few efforts have focused on developing tools for detecting misconfigurations present in DNS zone files. Pappas *et al.* [32] developed a tool to detect certain errors and inconsistencies by considering measurements from many vantage points. Many other tools are available online, for example at dns.net [33]. These tools check for a variety of problems, including lame delegation, presence of addresses in private ranges, absence of a prescribed number of name servers, invalid SOA timer values, lack of MX records for the domain, and several others. These tools analyze a single zone at a time and are not designed for the type of Internet-wide analysis we perform in this study. However, they are useful for administrators who wish to find and correct the errors in their own zones.

DNS performance has also been measured from other perspectives. Fujiware *et al.* [34] look from the clients' perspective, examining the impact that misconfigurations in authoritative DNS servers can have on resolvers. Danzig *et al.* [35] and Brownlee *et al.* [36] passively measured one of the DNS root servers to determine the characteristics of traffic received.

## IX. Conclusion

In this paper, we investigated the diversity of Internet zones and deployment levels of various DNS-based technologies. We also studied the intertwined relationships embedded in the various DNS records and their implications on availability of servers. The Internet-wide nature of our analysis allowed us to understand the common configuration mistakes that administrators make. While we found many distinct configuration problems, most were not very widespread, and not all were directly harmful to DNS operation. Administrators should however be careful to properly configure contact information, and to consider what information about their networks they are exposing to the outside world.

This study provides a snapshot in time of technology deployment and configuration problems in the DNS. Such deployments and configuration problems are likely to change over time. Due to the issues mentioned in Section III-A, and because we expect the availability of zone transfers to

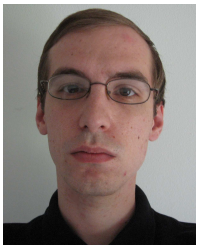
decrease, a comparison using similar methodology would be difficult in the future. However, at least some of the technologies and problems discussed in this paper may be tracked though other means. This paper provides a snapshot such future work can compare against to determine how much things have changed.

## ACKNOWLEDGMENTS

Rob Henderson helped in running the zone transfers smoothly. Without his help, it would not have been possible to collect data to conduct this research.

## REFERENCES

[1] R. Arends, R. Austein, M. Larson, D. Massey, and S. Rose, "Resource records for the DNS security extensions," IETF RFC 4034, Mar. 2005.
[2] A. Householder, B. King, and K. Silva, "Securing an internet name server," CERT Coordination Center Whitepaper, 2002.
[3] P. Mockapetris, "Domain names - concepts and facilities," IETF RFC 1034, Nov. 1987.
[4] ——, "Domain names - implementation and specification," IETF RFC 1035, Nov. 1987.
[5] VeriSign, Inc., "COM NET Registry TLD zone access program," http://www.verisign.com/information-services/naming-services/com-net-registry/page_001052.html.
[6] VeriSign, "Domain name industry brief," Jun. 2007, http://www.verisign.com/static/042161.pdf.
[7] O. Kolkman, M. Fuhr, D. Franks, and C. Reinhardt, "NET::DNS perl DNS resolver module," http://www.net-dns.org.
[8] E. Osterweil, D. Massey, and L. Zhang, "SecSpider," http://secspider.cs.ucla.edu.
[9] S. Josefsson, "DNSSEC walker," http://josefsson.org/walker/.
[10] SpamSuite.com, "Findings of fact, conclusions of law, and order for judgment," http://www.spamsuite.com/node/351.
[11] D. Eastlake 3rd, E. Brunner-Williams, and B. Manning, "Domain name system (DNS) IANA considerations," IETF RFC 2929, Sep. 2000.
[12] M. Lottor, "Domain administrators operations guide," IETF RFC 1033, Nov. 1987.
[13] D. Barr, "Common DNS operational and configuration errors," IETF RFC 1912, Feb. 1996.
[14] U. of Oregon Advanced Network Technology Center, "Route Views project," http://www.routeviews.org/.
[15] E. Osterweil, D. Massey, and L. Zhang, "Observations from the DNSSEC deployment," in *Workshop on Secure Network Protocols (NPSec)*, 2007.
[16] R. Ardnds, R. Austein, M. Larson, D. Massey, and S. Rose, "Protocol modifications for the DNS security extensions," IETF RFC 4035, Mar. 2005.
[17] C. Shue and M. Gupta, "Projecting IPv6 forwarding characteristics under Internet-wide deployment," in *ACM SIGCOMM 2007 IPv6 Workshop*, Aug. 2007.
[18] R. Hinden, S. Deering, and E. Nordmark, "IPv6 global unicast address format," IETF RFC 3587, Aug. 2003.
[19] J. Schlyter and W. Griffin, "Using DNS to securely publish secure shell (SSH) key fingerprints," IETF RFC 4255, Jan. 2006.
[20] M. Delany, "Domain-based email authentication using public keys advertised in the DNS (DomainKeys)," IETF RFC 4870, May 2007.
[21] E. Allman, J. Callas, M. Delany, M. Libbey, J. Fenton, and M. Thomas, "DomainKeys identified mail (DKIM) signatures," IETF RFC 4871, May 2007.
[22] J. Lyon and M. W. Wong, "Sender ID: Authenticating e-mail," IETF RFC 4406, April 2006.
[23] M. W. Wong and W. Schlitt, "Sender policy framework (SPF) for authorizing use of domains in e-mail, version 1," IETF RFC 4408, Apr. 2006.
[24] J. Rosenberg and H. Schulzrinne, "Session initiation protocol (SIP): Locating SIP servers," IETF RFC 3263, Jun. 2002.
[25] R. Elz, R. Bush, S. Bradner, and M. Patton, "Selection and operation of secondary DNS servers," IETF RFC 2182, Jul. 1997.
[26] Internet Assigned Numbers Authority, "Machine Names," http://www.iana.org/assignments/machine-names.
[27] ——, "Operating System Names," http://www.iana.org/assignments/operating-system-names.
[28] W. van Wanrooij and A. Pras, "DNS zones revisited," in *Open European Summer School and IFIP WG6.4/6.6/6.9 Workshop (EUNICE)*, 2005.
[29] V. Pappas, Z. Xu, S. Lu, D. Massey, A. Terzis, and L. Zhang, "Impact of configuration errors on DNS robustness," *ACM SIGCOMM Computer Communications Review (CCR)*, vol. 34, no. 4, pp. 319–330, 2004.
[30] The Measurement Factory, "DNS survey: October 2007," http://dns.measurement-factory.com/surveys/200710.html.
[31] A. Kalafut, C. Shue, and M. Gupta, "Understanding implications of DNS zone provisioning," in *ACM SIGCOMM Internet Measurement Conference (IMC)*, 2008.
[32] V. Pappas, P. Fältström, D. Massey, and L. Zhang, "Distributed DNS troubleshooting," in *ACM SIGCOMM Workshop on Network Troubleshooting*, 2004.
[33] A. Salamon, "Tools to manage DNS," http://www.dns.net/dnsrd/tools.html.
[34] K. Fujiwara, K. Toyama, K. Ishibashi, and C. Yoshimura, "DNS authoritative server misconfiguration," IETF Internet Draft, Feb. 2005.
[35] P. Danzig, K. Obraczka, and A. Kumar, "An analysis of wide-area name server traffic," in *ACM SIGCOMM*, 1992.
[36] N. Brownlee, K. Claffy, and E. Nemeth, "DNS measurements at a root server," in *IEEE Global Telecommunications Conference (GLOBECOM)*, 2001.



**Andrew J. Kalafut** received his Ph.D in Computer Science from Indiana University in 2010. He is currently an Assistant Professor in the School of Computing and Information Systems at Grand Valley State University. His research interests are in network measurement and security.

**Craig A. Shue** received his Ph.D in Computer Science from Indiana University in 2009. He is currently a Cyber Security Research Scientist in the Cyberspace Sciences and Information Intelligence Research Group at Oak Ridge National Laboratory. His research interests are in online deception, measurements, and Web security.

**Minaxi Gupta** Minaxi Gupta is an Associate Professor in the School of Informatics and Computing at Indiana University (Bloomington). Her research interests are in Computer Networks and Security. Gupta holds a Ph.D. in Computer Science from Georgia Tech, which she received in 2004.