Geo-Spatial Autocorrelation as a Metric for the Detection of Fast-Flux Botnet Domains

Etienne Stalmans
Security and Networks Research Group
Department of Computer Science
Rhodes University
Grahamstown,South Africa
E-mail: g07s0924@campus.ru.ac.za

Samuel Oswald Hunter
Security and Networks Research Group
Department of Computer Science
Rhodes University
Grahamstown,South Africa
E-mail: sam@rootentropy.co.za

Barry Irwin
Security and Networks Research Group
Department of Computer Science
Rhodes University
Grahamstown,South Africa
E-mail: b.irwin@ru.ac.za

***Abstract*—Botnets consist of thousands of hosts infected with malware. Botnet owners communicate with these hosts using Command and Control (C2) servers. These C2 servers are usually infected hosts which the botnet owners do not have physical access to. For this reason botnets can be shut down by taking over or blocking the C2 servers. Botnet owners have employed numerous shutdown avoidance techniques. One of these techniques, DNS Fast-Flux, relies on rapidly changing address records. The addresses returned by the Fast-Flux DNS servers consist of geographically widely distributed hosts. The distributed nature of Fast-Flux botnets differs from legitimate domains, which tend to have geographically clustered server locations. This paper examines the use of spatial autocorrelation techniques based on the geographic distribution of domain servers to detect Fast-Flux domains. Moran's I and Geary's C are used to produce classifiers using multiple geographic co-ordinate systems to produce efficient and accurate results. It is shown how Fast-Flux domains can be detected reliably while only a small percentage of false positives are produced.**

***Index Terms*—Botnets, Fast-Flux, Spatial Statistics**

I. Introduction

The connected nature of the modern Internet has led to a variety of computing devices from all over the world being present and active on the Internet. These devices are used by both legitimate users as well as malicious users of the internet. Many devices on the network are susceptible to infection by malicious programs known as malware. Malware allows malicious individuals to remotely control computing devices, and perform criminal activities such as sending spam, Denial of Service attacks and phishing. A large collection of these infected computing devices are known as a botnet. Botnets are controlled by malicious users through a central server architecture, using Command and Control (C2) servers. At the time of writing the ShadowServer organisation was tracking 1650 botnets, each containing an average of 150,000 infected hosts [1].

To prevent the C2 servers from being easily shut down, and with them the botnet, botnet controllers rely on different techniques. One of these techniques is known as Fast-Flux. Fast-Flux relies on the Domain Name System (DNS), which maps domain names to numeric IP addresses. Domains using Fast-Flux return multiple rapidly changing IP addresses for C2 servers with each DNS query response. These servers normally consist of compromised hosts that are under the control of the botnet owners. The infected hosts range from home computer systems to corporate computer networks, and are located world-wide [2]. The distributed nature of these C2 servers differ from legitimate domain servers. Legitimate domain servers under the control of organisations are usually located in a central geographic locations or closely clustered in distributed cases. Furthermore, legitimate domains such as Content Distribution Networks, that do use geographically distributed servers display a degree of geographic intelligence, where users are directed to servers that are geographically close to them [3]. This distinguishing factor allows for the differentiation between legitimate DNS domain entries and Fast-Flux domain entries. Classifiers based on animal and plant dispersion statistics allow for the accurate differentiation between legitimate and Fast-Flux domains. The classifiers produced are both accurate and lightweight, with no additional network traffic being generated and no time delay between a domain being seen for the first time and its classification.

The paper discusses related work in Section II. The datasets used are described in Section III. The geographic distribution of Fast-Flux C2 servers is discussed along with geographic co-ordinate systems in Section IV. Spatial autocorrelation and subsequent classification models used are detailed and discussed in Section V. The results obtained from testing sample traffic are presented in Section VI. These results are discussed and classifier performance outlined in Section VII, with concluding remarks in Section VIII.

II. Related Work

The distributed nature of botnet C2's is a well established fact and numerous researchers have attempted detection and classification of botnets using the geographic locations of botnet nodes [2], [3], [4]. In [4] Huang, Mao and Lee proposed a method for delay-free detection of Fast-Flux service networks. The solution relied on spatial distribution estimation and spatial service relationship evaluation. Timezones were used to distinguish between different geographic system spaces and was combined with information entropy to measure how uniformly nodes were distributed. The authors noted that benign domains tend to be distributed in the same timezone,

while Fast-Flux nodes are widely distributed across multiple timezones. Huang *et al*. further noted if all the hosts of a botnet were to occur in the same timezone, timezone based entropy would not be an effective measure for detecting if the hosts belonged to a benign or Fast-Flux domain.

The work performed by Caglayan, Toothaker, Drapaeau, Burke and Eaton [2] aimed to model the behavioral patterns of Fast-Flux botnets. Using DNS records, they showed Fast-Flux botnets exhibit common characteristics: that botnets form clusters based on botnet size, growth and operation. Furthermore, it was shown that the majority of Fast-Flux botnets operate in more than five countries at a time, averaging between 20 and 40 countries. Hu, Knysz and Shin [3] studied the global IP usage patterns of Fast-Flux botnets. Their work benefited from having a global perspective, with 240 nodes on four continents monitoring DNS traffic. Hu *et al*. found that Fast-Flux botnets advertise IP addresses from multiple countries, irrespective of where the DNS query came from, where-as Content Distribution Networks (CDNs) advertise IPs in a geographically aware manner. This observation provides valuable insight into the operation of Fast-Flux botnets, and furthermore helps determine how a classifier that is capable of differentiating between Fast-Flux botnets and CDNs may be constructed.

This paper differs from previous works by taking the novel approach of applying statistical methods from animal and plant dispersion research. This novel technique allows for classification of domains, solely from the geographic dispersion of the associated nodes. Classification allows for botnet C2 distribution to be labeled as randomly distributed or clustered. In previous work by the authors, it was shown how Fast-Flux botnets could be accurately detected using only DNS queries [5]. This paper aims to build on this work and possibly lead to a fully DNS-based malware detection framework.

III. Dataset

The datasets used were divided into *malicious* (Fast-Flux) data and *legitimate* (safe) data. The legitimate domain data was obtained from the "Alexa Top 1000 Global Sites" list [6]. The Alexa dataset was cross-correlated with the "Google Doubleclick Ad Planner Top-1000 Most Visited Sites" list [7]. The malicious data was taken from multiple sources, including the Fast-Flux trackers for the ZeuS [8], SpyEye [9] and other Fast-Flux botnets [10]. Data for the Hlux2/Khelios botnet was obtained from a large European ISP. The ZeuS Fast-Flux domains were obtained before the Microsoft take-down of ZeuS domains in March 2012 [11].

IV. Geographic Dispersion

The Domain Name System is used to resolve multiple network addresses to a central domain name. Domain name lookups allow infected hosts in the botnet to look up the address of C2 servers from which they need to receive instructions. Nazario and Holz noted that hosts used as C2 servers for a botnet need to meet specific criteria. These criteria include a globally accessible, globally unique IP address [12]. In further work Holz, Christian, Konrad and Freiling [13] identified the inherent distributed structure of botnets as a distinguishing factor. To contrast, legitimate domains tend to be set up with geographic location in mind, with all servers for the domain hosted in a central location, such as a data-center.

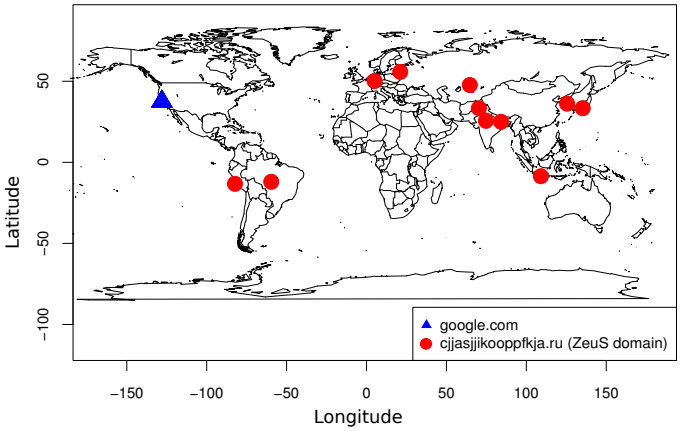


Figure 1: Geographic Distribution of Hosts for a Botnet and a Legitimate Domain

As seen in Figure 1 a Fast-Flux domain such as *cjjasjjikooppfkja.ru* has C2 servers widely dispersed globally, in this particular instance 11 different servers in 11 different countries across three continents. To contrast, a legitimate domain such as *google.com* has all six servers returned by a DNS query result located in one central location. The timezone in which a server is located is also of use as noted by Huang *et al.* [4]. The geographic distribution can be further analysed on a more finer grained level than just by country level, using co-ordinate systems such as the Universal Transverse Mercator system (UTM) and the Military Grid Reference System (MGRS). Table I provides the translation of IP addresses to geographic locations for a known Fast-Flux domain using the three different co-ordinate systems.

Table I: Geographic Data for a Fast-Flux Domain (*cjjasjjikooppfkja.ru*)

IP Address	Latitude:Longitude	UTM	MGRS
79.108.149.71	38.25:-0.7	37M	30SYH0125936055
79.139.110.20	49.7833:22.7833	39Q	34UFA2837416063
31.45.148.102	38.0:-97.0	37Z	14SPH7560307702
88.132.63.164	47.0333:19.7833	38Q	34TDT0755809583
124.6.3.225	22.6333:120.35	34Z	51QTF2762705352
89.229.214.126	53.7333:18.9167	39Q	34UCE6257855864

A. GeoIP Database

The company MaxMind has developed an IP Intelligence database [14], that contains geographic information for IP addresses throughout the entire world. The GeoIP City database, used by this research, allows for the country, city, latitude, longitude and other information pertaining to an IP address to be retrieved.

B. Geographic Value

The geographic locations for each server needed to be assigned a numerical value to be used in calculations. These values were obtained using three different means. Table II shows the numerical values as calculated from a servers timezone, UTM grid location and MGRS grid location.

Table II: Numeric Values for the IP Address 59.146.177.153

Method	Actual Value	Numeric Value
Latitude:Longitude	35.685°:139.7514°	-
Timezone	Asia/Tokyo	1000
UTM	36Z	3240
MGRS	54SUE8701849729	78943180741488

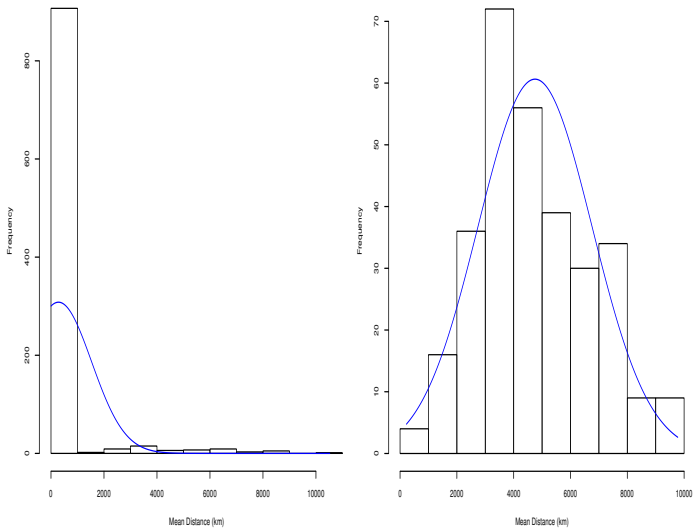
1) Timezone: The timezone in which a server is located allows for a value that is easy to calculate and quantify. Using Greenwich Meridian Time (GMT) as the base value of zero, each timezone was assigned a value. GMT+1 was assigned the value of 100, GMT+2 the value 200 and so forth. This was repeated for the timezones GMT-1 (100),GMT-2 (200).

2) Universal Transverse Mercator Coordinate System : The Universal Transverse Mercator (UTM) coordinate system is an alternative to the standard latitude and longitude coordinate system. Developed for use by the American Army, UTM is based on an ellipsoidal model of the Earth. The UTM system allows the earth to be divided into sixty distinct zones, each zone representing a six-degree band of longitude. For UTM to be used in the calculations discussed in Section V the UTM value needed to be converted to a fully numeric value. This was achieved by multiplying the numeric grid designator with the ordinal value of the alphabetic grid designator.

3) Military Grid Reference System: The Military Grid Reference System (MGRS) was developed to standardise geocoordinatation between NATO militaries. The MGRS is based on the UTM grid system and the similar Universal Polar Stereographic grid system. MGRS allows coordinates to be described as a grid point, down to one square meter. A MGRS grid point is identified by a grid zone designation, followed by a 100,000-meter square identification. For example: using MGRS, the latitude (26.12) and longitude (28.04) for Johannesburg, South Africa can be represented as 35RPJ3997589726. The MGRS value provides a grid location for each C2 server. This grid location needs to be converted into a numeric value to be used with the calculations discussed in Section V. This was achieved by multiplying the value V1 with the value V2. V1 was calculated using the first numeric grid descriptor multiplied with the ordinal values of the alphabetic grid descriptors. V2 was the integer value of the 100,000-meter square identification.

V. Spatial Autocorrelation

Correlation is used to measure the dependence or statistical relationship between any two points in a distribution. This correlation can refer to any characteristic that the points share such as geographic location, value or dependence on other points. Autocorrelation refers to the cross-correlation of a signal by itself, allowing for noise to be removed from the distribution. In statistics, autocorrelation is used for finding repeating patterns in a distribution. This has lead to the use of autocorrelation in different fields of study such as signal processing. While autocorrelation measures the dependence of points in time, spatial autocorrelation was developed to measure the dependence of points in two-dimensional space. Spatial autocorrelation is a branch of statistics that measures the spatial dependence of points within a geographic area. This measure of dependence is based on the principle that observations in close proximity are more likely to be similar than observations that are distant [15]. Spatial autocorrelation has largely been used in animal population statistics and disease modeling [15]. It is hypothesised that the principles behind animal population statistics and distribution modeling can be applied to the geographic distribution patterns of Fast-Flux botnets. Figures 2 clearly show that legitimate domain servers (Figure 2a) tend to be closer together, with the majority of servers being in the same location, while the C2 servers of Fast-Flux domains (Figure 2b) tend to be far apart with a binomial distribution centered around a mean nearest neighbour distance of 5000km. The use of spatial autocorrelation reduces the effect large outliers have on the overall classification.



(a) Histogram for Legitimate Domains (b) Histogram for Fast-Flux Domains

Figure 2: Histograms of Mean Nearest Neighbour Distances for Legitimate Domains (2a) and Fast-Flux (2b)

A. Moran's I

Moran's coefficient, known as Moran's I, provides a measure of Spatial autocorrelation. Moran's I is based on the observation that points spatially closer together are more likely to be similar than points far apart [16]. Values returned using Moran's I range from -1 to +1. Negative spatial correlation is indicated by values less than zero, while positive spatial correlation is indicated by values greater than zero. An index value of zero represents a perfectly random spatial pattern. Values

outside the range -1 to +1 indicate spatial autocorrelation that is significant at the 5% level. Moran's I can be calculated using the formula:

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$$

Where:

- I is the Moran Index
- N is the number of locations returned by the DNS query;
- $X_n$is the value of the n^{th} variable of interest (timezone value, UTM value, MGRS value);
- $\bar{X}$ is the average of all values of N;
- w_{ij} is the weight (distance) between two spatial points i and j.

Moran's I allows for the measuring of global spatial autocorrelation and is less influenced by large amounts of whitespace, making it ideal for use in classifying Fast-Flux C2 server distribution as there are large distances between the servers. Inversely, the global nature of Moran's I decreases it's effectiveness for measuring localised spatial correlation.

B. Geary's C

Similarly to Moran's I, Geary's C is used to measure spatial autocorrelation. The value of Geary's C lies in the range 0 and 2. Values between 0 and 1 indicate positive spatial autocorrelation while values between 1 and 2 indicate negative spatial autocorrelation. A value of 1 for Geary's C indicate no spatial autocorrelation. Geary's C can be defined by:

$$C = \frac{(N-1)\sum_i \sum_j w_{ij}(X_i - X_j)^2}{2W\sum_i (X_i - \bar{X})^2}$$

Where:

- C is the Geary value
- N is the number of locations returned by the DNS query;
- $X_n$is the value of the n^{th} variable of interest (timezone value, UTM value, MGRS value);
- $\bar{X}$ is the average of all values of N;
- w_{ij} is the weight (distance) between two spatial points i and j;
- W is the sum of all w_{ij}.

VI. Results

The aim of this research was to classify domains as either Fast-Flux or legitimate, based on the the geographic distribution of the domains servers. Different methods of indicating geographic position were used, including the Latitude/Longitude, the UTM grid position and the MGRS grid position. Distance between geographic points were measured using the Haversine formula which measures the distance between two points on a curved surface.

A. Moran's I

Moran's I relies on the observation that points closer together in geographic space tend to have more similarities than points far apart. All calculations were performed using the formula for Moran's I outlined above. The values for I were calculated separately for legitimate and Fast-Flux domains. Once these values had been calculated they were compared to see if there was a distinguishing value which could be used for accurately classifying the domains.

Table III: Moran's I Classifier Performance Using Different Geographic Properties

	TPR	FPR	ACC
Timezones	96.73%	3.42%	96.61%
UTM	99.03%	6.12%	95.35%
MGRS	99.35%	6.02%	95.28%

TPR: True Positive Rate, FPR: False Positive Rate, ACC: Accuracy

1) Timezones: The index value returned by the Moran formula, *I*, was used to classify a domain as legitimate or Fast-Flux. Results showed that the index value for legitimate domains was generally zero, with 97% of observed legitimate domains having an index value of zero. The opposite holds true for Fast-Flux domains, with only 3% of observed domains having an index value of zero. Using this as a classifier, domains were labeled as Fast-Flux if the returned index value was not equal to zero. Results for this classifier can be seen in Table III, where the classifier has a high true positive rate of 96.73% with a low false positive rate of 3.42%.

2) Universal Transverse Mercator (UTM): Using the UTM grid location of a server provides a more fine grained result than timezones as they cover a smaller area and take into account the hemisphere in which a server is located. Figure 3 compares the density distribution of the Moran index value for legitimate (Alexa Top 1000) domains and Fast-Flux domains. It is observed that the index value for legitimate domains is zero, or very near zero, while Fast-Flux domains tend to have an index value of one or greater. Based on a classifier value of zero where any index value greater than zero indicates a Fast-Flux domain, a true positive rate of 99.03% was achieved as seen in Table III, while a low false positive rate of 6.12% was achieved. This resulted in an overall accuracy of 95.35% for the classifier.

3) Military Grid Reference System (MGRS): The MGRS provides a grid system with smaller grids than the UTM coordinate system. Using MGRS produced interesting results as the index value for Fast-Flux domains was distributed between -0.5 and zero. While the index value for legitimate domains was mostly zero, a smaller cluster formed around an index value of -1. Basing the classifier on the same logic as was used for the timezone and UTM classifier, with an index value of zero indicating a legitimate domain, a true positive rate of 99.35% was achieved. A lower false positive rate of 6.02% was achieved. As seen in Figure 4, the Moran's I for numerous legitimate domains is -1, with no Fast-Flux domains having a

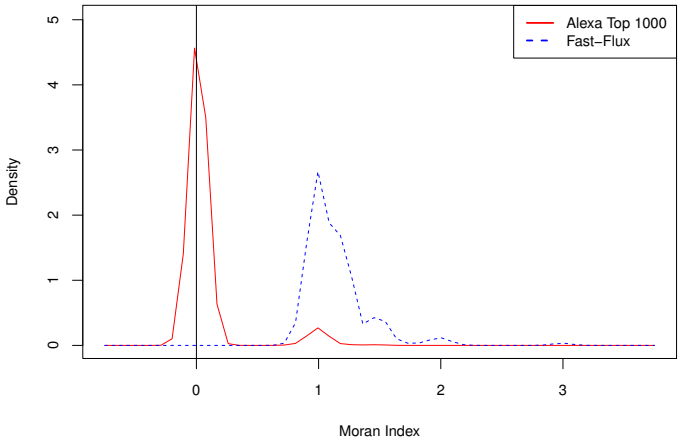


Figure 3: Kernel Density Comparison of Moran's I for Legitimate and Fast-Flux Domains Using UTM

Moran's I of -1. Thus modifying the classifier by classifying any value of -1 or 0 as legitimate led to an improved false positive rate of 1.24%, increasing the classifiers accuracy to 98.89%.

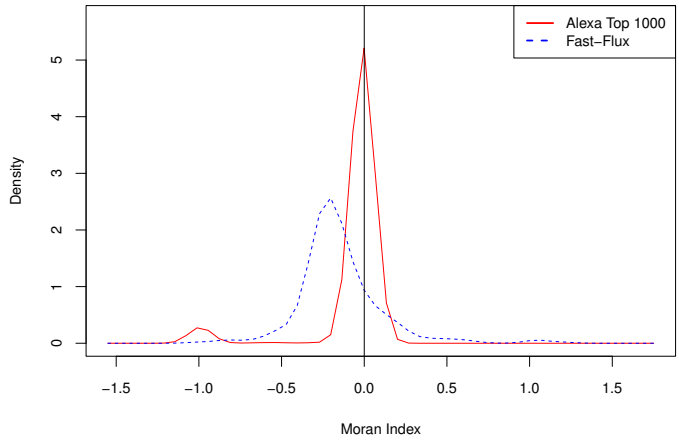


Figure 4: Kernel Density Comparison of Moran's I for Legitimate and Fast-Flux Domains Using MGRS

For simplicity and continuity, the results reported in Table III were based on the same classifier value of zero as was used for the timezones and UTM methods.

B. Geary's C

Geary's C is used for spatial autocorrelation, but is more sensitive to localisation than Moran's I. The use of Geary's C should allow for classification of spatial clusters in instances where Moran's I might not be as accurate. This might occur when the geographic location of servers are close together.

1) Timezones: The value produced by Geary's formula was used as the basis for classifying a domain as legitimate or Fast-Flux. The Geary C value was returned as zero for 96.58% of all legitimate domains, while returning a value greater than zero for 92.3% of Fast-Flux domains. Using this as the classification criteria, a classifier accuracy of 95.43% was achieved with a true positive rate of 91.83% and a false positive rate of 3.42%, as seen in Table IV.

Table IV: Geary's C Classifier Performance Using Different Geographic Properties

	TPR	FPR	ACC
Timezones	91.83%	3.42%	95.43%
UTM	98.37%	4.56%	96.14%
MGRS	99.64%	6.02%	95.35%

TPR: True Positive Rate,FPR: False Positive Rate,ACC: Accuracy

2) Universal Transverse Mercator (UTM): As seen with the Moran's I (Section VI-A2), UTM provides a higher degree of certainty when classifying domains due to the more fine grained nature. The Geary's C value for legitimate domains was clustered around zero as with the Moran's I. The value for Fast-Flux domains displayed two clusters, one at 0.5 and a second around 1. This led to the use of the same classifying criteria as before, with a value of zero indicating a legitimate domain while a value greater than zero indicated a Fast-Flux domain. The results from this classifier resulted in a 98.37% true positive rate with a low false positive rate of 4.56%. The overall accuracy for this classifier was a high 96.14%.

3) Military Grid Reference System (MGRS): The expected results from the MGRS classifier were a greater degree of accuracy than the UTM classifier for both Moran's I and Geary's C. Values returned for legitimate domains were clustered at zero, as seen in Figure 5. The values of C for Fast-Flux domains were clustered around 1. There were no negative values for Geary's C as expected (see Section V-B). The classifier was constructed with a value of zero indicating a legitimate domain and any value above zero indicating a Fast-Flux domain. The performance of the classifier was in line with the performance of the MGRS classifier based on Moran's I, with a similar true positive rate of 99.64%. The false positive rate of the classifier remained low at 6.02% and resulted in a classifier accuracy of 95.35%.

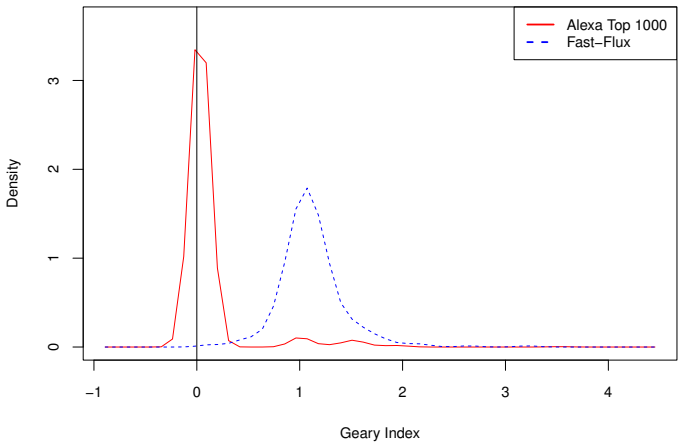
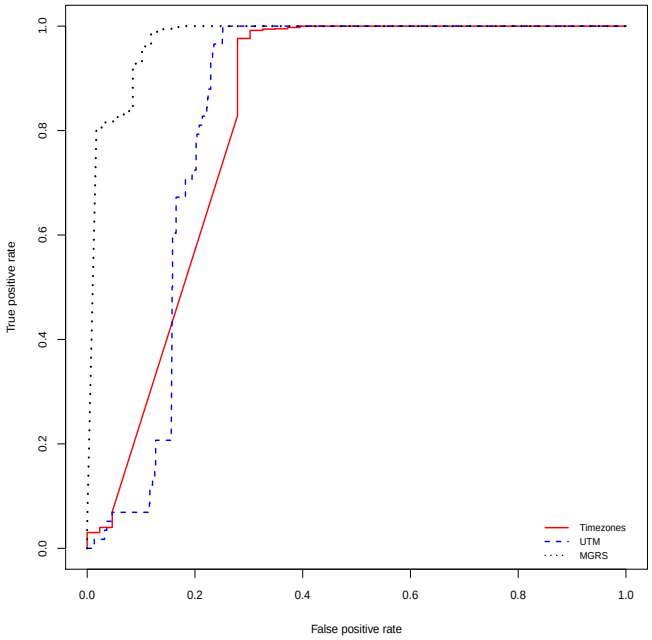
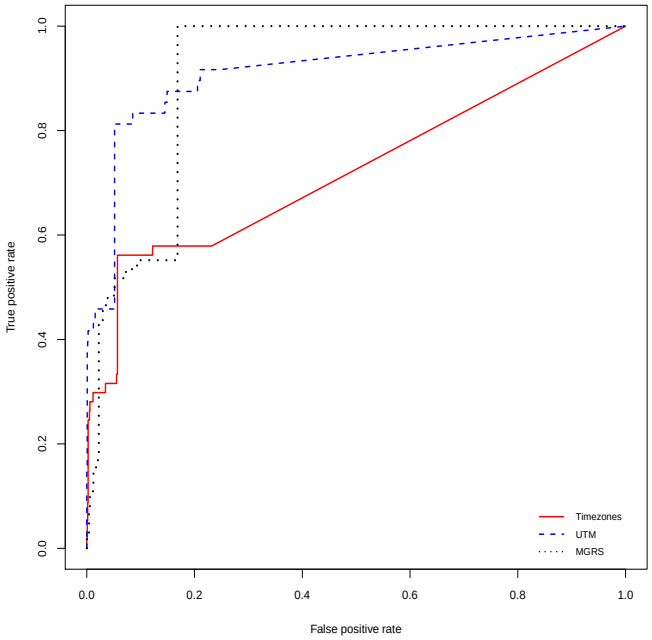


Figure 5: Kernel Density Comparison of Geary's C for Legitimate Domains and Fast-Flux Domains



(a) ROC for Moran's I Using Different Geolocation Features



(b) ROC for Geary's C Using Different Geolocation Features

Figure 6: ROC for Moran's I (6a) and Geary's C (6b)

VII. Discussion

The performance of Moran's I and Geary's C are closely matched and provide high rates of accuracy when trying to distinguish between legitimate and Fast-Flux domains based on the geographic locations of the domain servers. Figure 6 compares the performance of the six different classifiers. Figure 6a compares the performances of the three classifiers based on Moran's I, while Figure 6b compares the performances of the classifiers based on Geary's C. Using the area under the curve (AUC) as a performance metric, the classifier based on Moran's I, using the MGRS locations as values, produced the most efficient classifier, with an AUC of 98.54%. The least effective classifier is the Geary's C based classifier using the timezone values, with an AUC of 72.30%. The AUC for all classifiers is seen in Table V.

Table V: Area Under Curve Comparison for Multiple Classifiers

Geographic Feature	Moran's I	Geary's C
Timezone	83.50%	72.30%
UTM	83.60%	91.23%
MGRS	98.54%	91.10%

Each classifier employed used a value of zero as an identifier of legitimate domains, while any other value was classified as Fast-Flux. The short distance between legitimate server locations has been identified as a possible reason for legitimate queries returning a Moran's I or Geary's C of zero. Legitimate domains tend to have servers located in the same geographic location, while legitimate servers that do have servers in different geographic locations tend to have a mean nearest neighbour distance of 1000km. Fast-Flux C2 servers tend to be widely distributed, with no two servers located in the same location. The mean nearest neighbour distance for these domains is 5000km. It was also observed that legitimate domain queries tend to return fewer servers per DNS query while Fast-Flux domains returned a large number of C2 servers per query. Responses with a single location will have a Moran's I and Geary's C of zero, while the greater the number of locations returned, the more likely a larger value for Moran's I and Geary's C. It was observed that using a modified classifier value for the MGRS based Moran's I classifier could lead to greatly increased accuracy with a low false positive rate. This observation may be applied to future work where a learning algorithm could be used for determining the optimum classifier value.

The expected outcome was that Geary's C would be a more effective measure than Moran's I as it handles localized clustering better. This did not hold true, as seen from Table V, the performance of Moran's I was better than Geary's C in certain circumstances. The overall performance difference of the two classifiers was, however, minimal and this may be attributed to the relatively small number of locations tested on each occasion. It is expected that clustering of several hundreds or thousands of different locations will be better handled by Geary's C where as Moran's I will become affected

by the large amount of whitespace on a global scale.

VIII. Conclusion

Fast-Flux domains use multiple different servers located around the world to provide robust botnet control systems that are more resilient to shutdown attempts. This has lead to the development of systems that are capable of identifying Fast-Flux domains and to prevent access to those domains. This paper examined techniques for detecting Fast-Flux Command and Control (C2) domains. The techniques examined were able to accurately identify Fast-Flux domains based solely on the geographic locations of the C2 servers. The use of spatial autocorrelation has lead to the creation of accurate and lightweight classifiers. Analysis was performed on known Fast-Flux domains, with the most accurate classifier correctly identifying 99.35% of domains as Fast-Flux. The proposed solution provides an accurate means for improving network egress filtering, furthermore providing an effective additional layer of network defense usable in conjunction with existing defense systems. Future work will expand on the use of geographic location analysis, combining geographic location with other features identified as unique to Fast-Flux domains.

References

[1] S. Organisation. (2012, February) Botnet statistics. [Online]. Available: http://www.shadowserver.org/wiki/pmwiki.php/Stats/BotnetCharts

[2] A. Caglayan, M. Toothaker, D. Drapaeau, D. Burke, and G. Eaton, "Behavioral patterns of fast flux service networks," in *Proceedings of the 2010 43rd Hawaii International Conference on System Sciences*, ser. HICSS '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 1–9.

[3] X. Hu, M. Knysz, and K. G. Shin, "Measurement and analysis of global ip-usage patterns of fast-flux botnets." in *INFOCOM*. IEEE, 2011, pp. 2633–2641.

[4] S.-Y. Huang, C.-H. Mao, and H.-M. Lee, "Fast-flux service network detection based on spatial snapshot mechanism for delay-free detection." in *ASIACCS*, D. Feng, D. A. Basin, and P. Liu, Eds. ACM, 2010, pp. 101–111.

[5] E. Stalmans and B. Irwin, "A framework for dns based detection and mitigation of malware infections on a network," in *Proceedings of 9th Information Security South Africa (ISSA) Conference*, August 2011.

[6] Alexa. (2012, March) Alexa top sites. [Online]. Available: http://www.alexa.com/topsites

[7] Google. (2012) Top 1000 sites - doubleclick ad planner. [Online]. Available: http://www.google.com/adplanner/static/top1000/

[8] abuse.ch. (2012, February) Zeus monitor. [Online]. Available: https://zeustracker.abuse.ch/monitor.php?filter=level5

[9] ——. (2012) Spyeye monitor. [Online]. Available: https://spyeyetracker.abuse.ch/monitor.php?filter=level5

[10] ArborNetworks. (2012, March) Atlas global fast flux report. [Online]. Available: https://atlas.arbor.net/summary/fastflux

[11] Microsoft Corporation. (2012, March) Microsoft and financial services industry leaders target cybercriminal operations from zeus botnets. [Online]. Available: https://blogs.technet.com/b/microsoft_blog/archive/2012/03/25/microsoft-and-financial-services-industry-leaders-target-cybercriminal-operations-from-zeus-botnets.aspx

[12] J. Nazario and T. Holz, "As the net churns: Fast-flux botnet observations," in *2008 3rd International Conference on Malicious and Unwanted Software (MALWARE)*. IEEE, Oct. 2008, pp. 24–31.

[13] T. Holz, C. Gorecki, K. Rieck, and F. C. Freiling, "Measuring and detecting fast-flux service networks," in *MALWARE 2008. 3rd International Conference on Malicious and Unwanted Software, 2008*, 2008, pp. 24 – 31.

[14] MaxMind. (2012, February) Maxmind geoip. [Online]. Available: http://www.maxmind.com/app/ip-location

[15] O. Schabenberger and C. A. Gotway, *Statistical methods for spatial data analysis*, B. P. Carlin, C. Chatfield, M. Tanner, and J. Zidek, Eds. Chapman & Hall/CRC, 2005, vol. 64.

[16] A. Cliff and J. K. Ord, *Spatial Autocorrelation*. Pion, 1973.