On some information geometric approaches to cyber security–*Summary*

CTJ Dodson

School of Mathematics, University of Manchester, Manchester M13 9PL, UK
ctdodson@manchester.ac.uk

Abstract. Various contexts of relevance to cyber security involve the analysis of data that has a statistical character and in some cases the extraction of particular features from datasets of fitted distributions or empirical frequency distributions. Such statistics, for example, may be collected in the automated monitoring of IP-related data during accessing or attempted accessing of web-based resources, or may be triggered through an alert for suspected cyber attacks. Information geometry provides a Riemannian geometric framework in which to study smoothly parametrized families of probability density functions, thereby allowing the use of geometric tools to study statistical features of processes and possibly the representation of features that are associated with attacks. In particular, we can obtain mutual distances among members of the family from a collection of datasets, allowing for example measures of departures from Poisson random or uniformity, and discrimination between nearby distributions. Moreover, this allows the representation of large numbers of datasets in a way that respects any topological features in the frequency data and reveals subgroupings in the datasets using dimensionality reduction. Here some results are reported on statistical and information geometric studies concerning pseudorandom sequences, encryption-decryption timing analyses, comparisons of nearby signal distributions and departure from uniformity for evaluating obscuring techniques.

Keywords: cyber security, empirical frequency distributions, pseudorandom sequences, encryption-decryption timing, proximity to uniformity, nearby signals discrimination, information geometry, gamma distributions, Gaussian distributions, dimensionality reduction

1 Introduction

The British Columbia Institute of Technology (BCIT) maintained until 2006 an industrial cyber security incident database (ISID) [8], designed to track incidents of a cyber security nature that directly affected industrial control systems and processes. Byres and Lowe [9,8] pointed out that from 1980 to 2000 the cyber threat was evenly split among internal, external and accidental cases. By 2001 this had changed to 70% external threat sources, 20% accidental, 5% internal and 5% other. Of these the internal security incidents arose from the following entry points: business network 43%, human machine interface (HMI) 29%, physical access to equipment 21% and laptop 7%. Externally the percentages included attacks from: remote internet 36%, remote dial-up 20%, remote unknown 12%, VPN connection 8%, remote wireless 8%, the remainder from remote trusted third party, remote Telco network, remote supervisory control and data acquisition (SCADA) network. The consequences of these attacks was a production loss of 41% and loss of ability to control or view the plant. Between 1995 and 2000 the number of security incidents averaged 2 per year but that had increased linearly to 10 per year by 2003. The current successor to ISID is the Repository of Industrial Security Incidents [57], a database of incidents of a cyber security nature that have (or could have) affected process control, industrial automation or supervisory control and data acquisition (SCADA) systems. For a current view of the problem of criminal use of encrypted messaging systems on smartphones, see the New York District Attorney's report to the 6^{th} Annual Financial Crimes and Cybersecurity Symposium at the Federal Reserve Bank of New York on 15 November 2015 [62], with a large bibliography. This sets out the current capabilities of smartphones and tablets and makes a number of proposals.

The UK government Centre for the Protection of National Infrastructure (CPNI) [19] and the USA Homeland Security [37] provide up to date information and advice on cyber security. Wang and Lu [63] provided a comprehensive study of cyber security needs for the next generation power systems, particularly network vulnerabilities, attack countermeasures, secure communication protocols and architectures in the Smart Grid. The UK Information Assurance Advisory Council (IAAC) [52] provides a wide range of documentation, including the latest Korea-UK Initiatives in Cyber Security Research report [61]. The proceedings of the recent international conference at the University of Piraeus [45], provides a collection of more than 30 articles on cyber warfare and security and the book [56] contains 17 articles treating various aspects of cybersecurity. Via the assistance of the 2014 US AMS Network Science Mathematical Research Community, Burstein et al [7] studied the problem of increasingly frequent events of Border Gateway Protocol route hijacking for traffic interception. They developed an optimal information monitoring strategy based on an abstract model for routing networks in which colluding sets of agent nodes conspire to divert traffic via them by sending false distance information to honest agent nodes.

In this paper we offer some geometrical methods for application in problems of cyber security which can be addressed through statistical analyses of data. Information geometry provides a Riemannian geometric framework in which to study smoothly parametrized families of probability density functions, thereby allowing the use of geometric tools to study statistical features of processes. Geometrical provision of this kind has proved an enormous advantage in theoretical physics and conversely, physical problems have stimulated many advances in differential geometry, global analysis and algebraic geometry. The geometrization of statistical theory [1,2,3,4,5,22] has had similar success and its role in applications is now widespread and generating new developments of theory, algorithms and computational information geometry [50,51]. We give a brief introduction to information geometry in §2 and §2.1, which is sufficient for the understanding of techniques in the sequel. We outline the information geometry of univariate and multivariate Gaussians in §2.2, which we use in §7. Situations in which such methods are relevant to cyber security include discrimination between nearby signal distributions, comparisons of real signal distributions with those obtained via random number generators in testing obscuring procedures, and in testing for anomalous behaviour, for example using departures from uniformity or independence.

One aspect of cyber security is concerned with the analysis of the stochastic process of attack events [21]. Such analyses can yield valuable data on the frequency distributions of attacks and these may be amenable to study using information geometric methods. In particular, spacings between events of interest may be representable via gamma distributions, since they span a range of behaviour from clustered through random (ie Poisson) to dispersed, Figure 1; we discuss their information geometry in §3. Gamma distributions have the property that the standard deviation is proportional to the mean, characterized in Theorem **??** below, and they include a representation of Poisson processes through the 1-parameter family of exponential distributions; this is represented in Figure **??**. Their information geometry was used in a variety of applications [5,24]. In a range of contexts in cryptology for encoding, decoding or for obscuring procedures, sequences of pseudorandom numbers are generated. Tests for randomness of such sequences have been studied extensively and the NIST Suite of tests [55] for cryptological purposes is widely employed. Information theoretic methods also are used, for example see Czregorzewski and Wieczorkowski [20]

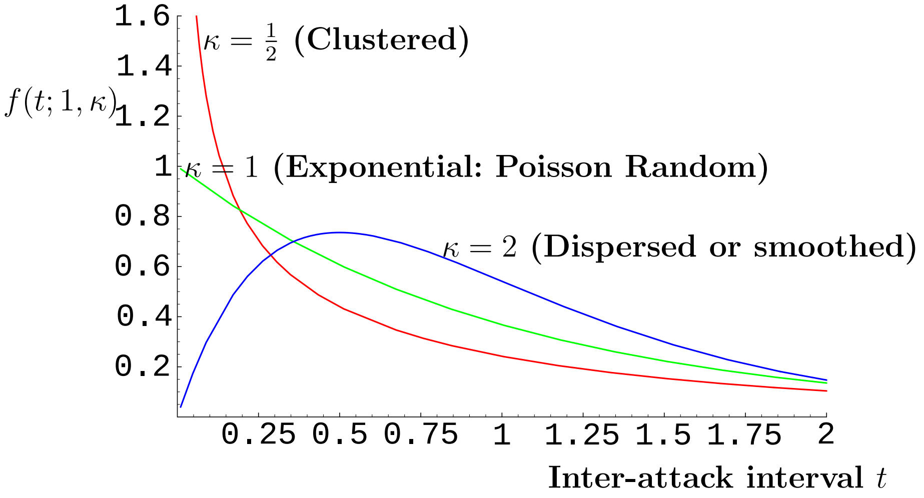

Fig. 1. *Probability density functions, $f(t;\mu,\kappa)$, for gamma distributions of inter-attack intervals t with unit mean $\mu = 1$, and $\kappa = \frac{1}{2}$, 1, 2. The case $\kappa = 1$ corresponds to an exponential distribution from an underlying Poisson process; $\kappa \neq 1$ represents some organization—clustering or dispersion.*

also Ryabko and Monarev [58] and references therein for recent work. Covert timing channels operate by establishing an illegitimate communication channel between two processes and transmitting information via timing modulation, violating the underlying system's security policy. Recent studies have shown the vulnerability of popular computing environments, such as cloud, to these covert timing channels. Chen and Venkataramani [18] proposed an algorithm to detect the possible presence of covert timing channels on shared hardware that use contention-based patterns for communication. They obtained an event density histogram to represent the probability distribution of event density and compared this to a Poisson process. We show in §4 how pseudorandom sequences may be tested using information geometry by using distances in the gamma manifold to compare maximum likelihood parameters for separation statistics of sequence elements.

In practical signal comparison situations [29], we obtain statistical data for an observable that is defined on some finite interval. We shall use as our model the family of log-gamma probability density functions, Figure **??**, defined for random variable $a \in (0, 1]$ in §3.1. The choice of log-gamma model is due to the fact that it contains a neighbourhood of the uniform distribution, and it has approximations to Gaussians truncated to domain $(0, 1]$ and with arbitrarily small variance. The role of these functions in testing we discuss in §5.

Encryption devices may be attacked by electromagnetic sensors that can extract information on the timing of processes for a chosen range of input data values. Given some knowledge of the software architecture, timings of operations typically relate to modular exponentiation steps, associated with the processing of the binary bits in the encryption key. This is discussed in §6. In practice, clues to such timing information can be obtained from data on power consumption using electromagnetic sensors, possibly needing statistical processes to clean the data of noise. Kocher et al [41] showed the effectiveness of Differential Power Analysis (DPA) in breaking encryption procedures using correlations between power consumption and data bit values during processing, claiming that most smart cards revealed their DES keys using fewer than 15 power traces. A practicable defence is to obscure the power usage data on timing information by spurious other processes. Then the effectiveness of such obscuring techniques can be evaluated using analyses of the distributions associated with time series from power usage. For example, a time series of power consumption using appropriately chosen threshholding and interval windows would yield a barchart and that would ideally be like that arising from Poisson processes, which for a given mean are maximally disorderd [36]. Information geometry can be used to measure differences from the Poisson model, equivalently from its associated exponential distribution—note that Grzegorzewski and Wieczorkowski [20] provided a detailed analysis of their entropy-based goodness-of-fit test for exponentiality.

Evaluation of cyber security may involve also identifying potentially anomalous behaviour in internet traffic on a network [53,48], thus requiring extraction of appropriate features from a large data set of event frequency distributions. sometimes we can fit standard models to the empirical frequency distributions using maximum likelihood methods as iillustrated for gamma distributions in §3. In the absence of a model family of distributions for which we have expressions for the information distances among the memberrs, we can use the symmetrized Kullback-Leibler relative entropy expression, equation (**??**), to measure distance between empirical frequency distributions. Once we have extracted distance measures between all pairs of datasets we can use multi-dimensional scaling, or dimensionality reduction, to extract the three most significant features from the data set so that all samples can be displayed graphically in a 3-dimensional plot. The aim is to reveal groupings of data points that correspond to the prominent characteristics, the methodology is discussed in §7.

Such a dimensionality reduction can reveal anomalous behaviour of a process by taking account of the true curved geometry of the data set, rather than displaying it as uncurved in a Euclidean geometry (cf. [12], Figure 3.2). The significance is that any non-obvious global topology of frequency connectivity in the data is revealed by the pattern of mutual separations in the embedding. An illustration using router traffic on the Abilene network showed how anomalous behaviour unseen by local methods could be picked up through dimensionality changes (cf. [12], Figure 3.10). Moreover, in document classification, the information metric approach outperformed standard Principal Component Analysis and Euclidean embeddings [13], and it outperformed traditional approaches to video indexing and retrieval with real world data [17]. In §7 we outline how autocovariance extraction from time series data may be studied using information geometry and dimensionality reduction; we described an application to datasets of stochastic textures from 2-dimensional pixel arrays in [28].

We begin here by outlining the method to compute the Fisher information metric on a smoothly parametrized family of probability density functions, then illustrate it with explicit expressions for some important examples.

2 The Fisher information metric

2.1 Exponential family of distributions

2.2 Information geometry of Gaussians

3 Information geometry of the gamma manifold

3.1 Neighbourhoods of uniformity in the log-gamma manifold

3.2 Neighbourhoods of randomness in the gamma manifold

4 Statistics of finite random spacing sequences

4.1 Derivation of the distributions

5 Testing nearby signal distributions and drifts from uniformity

6 Protecting devices with obscuring techniques

7 Dimensionality reduction methods

8 Discussion

References

1. S-I. Amari. Theory of Information Spaces—A Geometrical Foundation of the Analysis of Communication Systems. *Research Association of Applied Geometry Memoirs* 4 (1968) 171-216.
2. S-I. Amari. **Differential Geometrical Methods in Statistics** Springer Lecture Notes in Statistics 28, Springer-Verlag, Berlin 1985.
3. S-I. Amari, O.E. Barndorff-Nielsen, R.E. Kass, S.L. Lauritzen and C.R. Rao. **Differential Geometry in Statistical Inference**. Lecture Notes Monograph Series, Institute of Mathematical Statistics, Volume 10, Hayward California, 1987.
4. S-I. Amari and H. Nagaoka: *Methods of Information Geometry.* Oxford, American Mathematical Society, Oxford University Press (2000).
5. K. Arwini and C.T.J. Dodson: *Information Geometry Near Randomness and Near Independence.* Lecture Notes in Mathematics. New York, Berlin, Springer-Verlag (2008).
6. C. Atkinson and A.F.S. Mitchell. Rao's distance measure. *Sankhya: Indian Journal of Statistics* 48, A, 3 (1981) 345-365.
7. D. Burstein, F. Kenter, J. Kun and F. Shi. Information Monitoring in Routing Networks. (2015) 12 pages http://arxiv.org/pdf/1507.05206.pdf
8. E. Byrse and D. Leversage: The Industrial Security Incident Database. (2006)
http://www.securitymetrics.org/attachments/Metricon-1-Leversage-Rump.pdf
9. E. Byrse and J. Lowe: The Myths and Facts behind Cyber Security Risks for Industrial Control Systems. VDE 2004 Congress, VDE, Berlin, October 2004

10. B. Canvel. **Timing Tags for Exponentiations for RSA.** MSc Thesis, Department of Mathematics, University of Manchester Institute of Science and Technology, Manchester (1999).
11. B. Canvel and C.T.J. Dodson. *Public Key Cryptosystem Timing Analysis.* CRYPTO 2000, Rump Session Santa Barbara, 20-24 August 2000. `http://www.maths.manchester.ac.uk/~kd/PREPRINTS/rsatim.pdf 27 August 2000`
12. K.M. Carter. Dimensionality reduction on statistical manifolds. PhD thesis, University of Michigan, 2009. http://tbayes.eecs.umich.edu/kmcarter/thesis
13. K.M. Carter, R. Raich and A.O. Hero III. FINE: Information embedding for document classification. In Proc. 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing, Las Vegas, March 2008. `http://tbayes.eecs.umich.edu/kmcarter/fine_doc`
14. K.M. Carter, R. Raich,W.G. Finn and A.O. Hero III. Fisher information nonparametric embedding. *IEEE Trans. Pattern. Anal. Mach. Intell.* 31 (2009) 20932098. http://arxiv.org/abs/0802.2050v1
15. K.M. Carter, R. Raich,W.G. Finn and A.O. Hero III. Information-geometric dimensionality reduction. *IEEE Signal Processing Mag.* 99 (2011) 89-99. `http://web.eecs.umich.edu/~hero/Preprints/carter_spsmag_igdr_rev3.pdf`
16. S. Chari, C.S. Jutla, J.R. Rao and P. Rohatgi. Towards sound approaches to counteract power-analysis attacks. In **Advances in Cryptology-CRYPTO '99**, Ed. M. Wiener, Lecture Notes in Computer Science 1666, Springer, Berlin 1999 pp 398-412.
17. X. Chen and A. Hero. Fisher Information Embedding for Video Indexing and Retrieval. SPIE Electronic Imaging Conference, San Jose, 2011. `nts/ChenEI11.web.eecs.umich.edu/~hero/Prepripdf`
18. J. Chen and G. Venkataramani: An algorithm for detecting contention-based covert timing channels on shared hardware. In Proc. **HASP '14 Third Workshop on Hardware and Architectural Support for Security and Privacy** ACM Digital Library dl.acm.org `http://www.seas.gwu.edu/~guruv/hasp14.pdf`
19. CPNI: UK Centre for the Protection of National Infrastructure http://www.cpni.gov.uk/advice/cyber/
20. P. Crzegorzewski and R. Wieczorkowski. Entropy-based goodness-of-fit test for exponentiality. *Commun. Statist. Theory Meth.* 28, 5 (1999) 1183-1202.
21. N.J. Daras: Stochastic analysis of cyber attacks. In Applications of Mathematics and Informatics in Science and Engineering, Springer Optimization and its Applications 91, Ed. N.J. Daras, 105–129 (2014)
22. C.T.J. Dodson Editor. Proceedings of **Workshop on Geometrization of Statistical Theory** Lancaster 28-31 October 1987, ULDM Publications, University of Lancaster, 1987.
23. C.T.J. Dodson. Information distance estimation between mixtures of multivariate Gaussians. Presentation at Workshop on **Computational information geometry for image and signal processing** ICMS Edinburgh, 21-25 September 2015.
24. C.T.J. Dodson. Some illustrations of information geometry in biology and physics. In *Handbook of Research on Computational Science and Engineering: Theory and Practice* Eds. J. Leng, W. Sharrock, IGI-Global, Hershey, PA, 2012, pp 287-315. http://www.igi-global.com/book/handbook-research-computational-science-engineering/51940
25. C.T.J. Dodson and Hiroshi Matsuzoe. An affine embedding of the gamma manifold. *InterStat*, January 2002, 2 (2002) 1-6.

26. C.T.J. Dodson, M. Mettänen and W.W. Sampson: Dimensionality reduction for characterization of surface topographies. Presentation at Workshop on **Computational information geometry for image and signal processing** ICMS Edinburgh, 21-25 September 2015.
27. C.T.J. Dodson and T. Poston. **Tensor Geometry** Graduate Texts in Mathematics 130, Second edition, Springer-Verlag, New York, 1991.
28. C.T.J. Dodson and W.W. Sampson: Dimensionality reduction for classification of stochastic texture images In F. Nielsen(Ed.): **Geometric Theory of Information**. Springer, Heidelberg, 1013–1015 (2014)
29. C.T.J. Dodson and S.M. Thompson. A metric space of test distributions for DPA and SZK proofs. Poster Session, **Eurocrypt 2000**, Bruges, 14-19 May 2000. http://www.maths.manchester.ac.uk/~kd/PREPRINTS/mstd.pdf
30. P.S. Eriksen. Geodesics connected with the Fisher metric on the multivariate normal manifold. In C.T.J. Dodson, Editor, **Proceedings of the GST Workshop**, Lancaster (1987), 225-229. http://trove.nla.gov.au/version/21925860
31. W. Feller: *An Introduction to Probability Theory and its Applications.* Volume II 2^{nd} Edition, Wiley, New York (1971).
32. R.A. Fisher. Theory of statistical estimation. *Proc. Camb. Phil. Soc.* 122 (1925) 700-725.
33. P. Ginlin. **Primes and Programming: An Introduction to Number Theory with Computing.** Cambridge University Press, 1993.
34. O. Goldreich, A. Sahai and S. Vadham. Can Statistical Zero-Knowledge be made non-interactive? Or, on the relationship of SZK and NISZK. In **Advances in Cryptology-CRYPTO '99**, Ed. M. Wiener, Lecture Notes in Computer Science 1666, Springer, Berlin 1999 pp 467-484.
35. Y. Gu, A. McCallum and D. Towsley: Detecting anomalies in network traffic using maximum entropy estimation. In Proc. **Internet Measurement Conference 2005** pp 345-350. More details are in the Technical Report from the Department of Computer Science, UMASS, Amherst 2005.
36. F.A. Haight. **Handbook of the Poisson Distribution** J. Wiley, New York, 1967.
37. USA Homeland Security: Cybersecurity, http://www.dhs.gov/topic/cybersecurity 2015.
38. T-Y. Hwang and C-Y. Hu: On a characterization of the gamma distribution: The independence of the sample mean and the sample coefficient of variation. *Annals Institute Statistical Mathematics* 51(4) 749-753 (1999).
39. V. Jacobson, C. Leres and S. McCanne: *tcdump* via anonymous ftp.ee.lbl.gov, June 1989.
40. D.H. Johnson and S. Sinanovic: Symmetrizing the Kullback-Leibler Distance. *Rice University doc* https://scholarship.rice.edu/handle/1911/19969
41. P. Kocher, J. Jaffe and B.Jun. Differential Power Analysis. In **Advances in Cryptology-CRYPTO '99**, Ed. M. Wiener, Lecture Notes in Computer Science 1666, Springer, Berlin 1999 pp 388-397.
42. S. Kullback. **Information Theory and Statistics**. Wiley, New York, 1959.
43. S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Stat.*, 22 (1951) 79-86.
44. R.G. Laha: On a characterization of the gamma distribution. *Ann. Math. Stat.* 25 (1954) 784-787.
45. A. Liaropoulos and G. Tsihrintzis, Eds. **Proceedings of 13^{th} European Conference on Cyber Warfare and Security** University of Piraeus, Greece, 3-4 July 2014.

46. W. Lee and D. Xiang. Information-theoretic measures for anomaly detection. In Proc. **IEEE Symposium Security and Privacy** (2001) pp 130-143. 10.1109/SECPRI.2001.924294
47. Y. Liu: Intrusion detection for wireless networks. PhD thesis, Stevens Institute of Technology, ACM Digital Library dl.acm.org .
48. L.A. Maglaras, J. Jiang and T.J. Cruz. Integrated ocsvm mechanism for intrusion detection in scada systems. *Electronics Letters* 50, (2014) 19351936(1). Cf also: Combining ensemble methods and social network metrics for improving accuracy of OCSVM on intrusion detection in SCADA systems. http://arxiv.org/pdf/1507.02825.pdf , 25 pages.
49. E. Nash: Hackers bigger threat than rogue staff. VNU Publications, May 15, 2003.
50. F. Nielsen and F. Barbaresco (Eds.): **Geometric Science of Information** GSI2013, LNCS 8085, Springer, Heidelberg (2013).
51. F. Nielsen(Ed.): **Geometric Theory of Information**. Springer, Heidelberg (2014).
52. The Information Assurance Advisory Council (IAAC). http://www.iaac.org.uk/
53. Maxim Raginsky, Rebecca Willett, Corinne Horn, Jorge Silva and Roummel Marcia: Sequential anomaly detection in the presence of noise and limited feedback. *ArXiv* arXiv:0911.2904v4 (2012) 1-19.
54. C.R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* 37, (1945) 81-91.
55. A. Rushkin, J. Soto et al. **A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications**. *National Institute of Standards & Technology*, Gaithersburg, MD USA, 2001.
56. Robinson E. Pino, Ed. **Network Science and Cybersecurity**, Springer, New York, 2014.
57. RISI Online Incident Database. http://www.risidata.com/Database
58. B.Ya. Ryabko and V.A. Monarev. Using information theory approach to randomness testing. *J. Stat. Plan. Inf.* 133, 1 (2005) 95-110.
59. S.D. Silvey. **Statistical Inference** Chapman and Hall, Cambridge 1975.
60. SunSoft **SunSHIELD Basic Security Module Guide**. Soft, Mountain View, CA, 1995. https://docs.oracle.com/cd/E19457-01/801-6636/801-6636.pdf
61. Peter Trim and Heung Youl Youm, Eds. **Korea-UK Initiatives in Cyber Security Research: Government, University and Industry Collaboration**, Report Submitted to the Korean Government and the UK Government March, 2015.
http://www.iaac.org.uk/media/1356/cyber-security-report-trim-and-youm-march2015.pdf
62. Cyrus R. Vance Jr. Smartphone Encryption and Public Safety. 6th **Annual Financial Crimes and Cybersecurity Symposium** Federal Reserve Bank of New York 15 November 2015, 42 pages. http://manhattanda.org/sites/default/files/11.18.15
63. W. Wang and Z. Lu. Cyber security in the Smart Grid: Survey and challenges. *Computer Networks* 57, 5 (2013) 13441371.
http://www.sciencedirect.com/science/article/pii/S1389128613000042
64. S. Wolfram. **The Mathematica Book** 3rd edition, Cambridge University Press, Cambridge, 1996.