Front. Electr. Electron. Eng. China 2009, 4(3): 307–312
DOI 10.1007/s11460-009-0051-9

RESEARCH ARTICLE

Yanhua YU, Jun WANG, Xiaosu ZHAN, Junde SONG

Novel anomaly detection approach for telecommunication network proactive performance monitoring

Abstract The mode of telecommunication network management is changing from "network oriented" to "subscriber oriented". Aimed at enhancing subscribers' feeling, proactive performance monitoring (PPM) can enable a fast fault correction by detecting anomalies designating performance degradation. In this paper, a novel anomaly detection approach is the proposed taking advantage of time series prediction and the associated confidence interval based on multiplicative autoregressive integrated moving average (ARIMA). Furthermore, under the assumption that the training residual is a white noise process following a normal distribution, the associated confidence interval of prediction can be figured out under any given confidence degree $1-\alpha$ by constructing random variables satisfying t distribution. Experimental results verify the method's effectiveness.

Keywords proactive performance monitoring (PPM), anomaly detection, time series prediction, autoregressive integrated moving average (ARIMA), white noise, confidence interval

1 Introduction

The mode of telecommunication network management is changing from "network oriented" to "subscriber oriented". Accordingly, the mode of performance management of network, an important aspect of network management, is transforming from passiveness to initiativeness. The proactive performance monitoring (PPM) is simply one of the initiative management ways.

PPM [1,2] is concerned with performance analysis and fault detection, which are capable of detecting "soft" network and service faults automatically and adaptively in the midst of networks' performance fluctuations and evolutions. By detecting anomalies designating performance degradation, which are the symptoms of network faults and preludes to services failure, PPM can enable a fast fault containment and correction, through which serious network failures can be avoided and duration time can be shortened. "Anomaly" is defined as "statistically unusual". It can be detected by using two terms: baseline and thresholds (upper and lower thresholds). Thresholds fluctuate around predictive baseline. If the actual value of key performance indicator (KPI) exceeds thresholds, performance warning would be triggered to show the deteriorating performance quality.

For anomaly detection, the prediction of the baseline and thresholds of KPI is a crucial point. If the threshold is set over strictly, a fake warning may be launched. However, if the threshold is too high or too low, the mechanism will make little sense. There were some researches on the derivations of baseline and thresholds [2,3]. In Ref. [2], the threshold that is called tolerance limits or envelope is based on standard deviation of training set, while in Ref. [3], the baseline is based on a simplistic algorithm, and they are all too coarse. In this paper, time series prediction is deeply studied to settle baseline, and confidence interval of prediction error is used for thresholds. Under the assumption that a white noise process follows normal distribution, the associated confidence interval of prediction value can be worked out under any dedicated confidence degree $1-\alpha$ by constructing random variables following t distribution.

This paper is organized as follows. In Sect. 2, the principle and model identification method of autoregressive integrated moving average (ARIMA) is introduced. In Sect. 3, the proposed time series prediction approach with the associated confidence interval calculation using multiplicative ARIMA is presented in particular. In Sect. 4, an experiment with the proposed approach is carried out. The conclusion is drawn in Sect. 5.

Received October 16, 2008; accepted November 10, 2008

Yanhua YU (✉), Jun WANG, Xiaosu ZHAN, Junde SONG
School of Computer Science and Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
E-mail: yhyu_bupt@sina.com.cn

2 Overview of ARIMA

Let time series $x_1,x_2,\ldots,x_i,\ldots,x_N$ denote the observations made at equidistant time interval $\tau_0+h,\tau_0+2h,\ldots,\tau_0+ih,\ldots,\tau_0+Nh$, where x_i represents observation at time τ_0+ih if we consider τ_0 as the origin and h as the unit of time. Basically, the time series prediction can be considered as a modeling issue; a model is built between input and output. Then, the model is used to predict the future values based on the previous values.

Among the algorithms deployed in time series prediction, both autoregressive moving average (ARMA) [4] and ARIMA [5] are most commonly used. ARMA can only be used for linear and stationary time series, while ARIMA can be used for nonstationary ones exhibiting homogeneity. When the time series show heterogeneous nonstationarity, other algorithms such as artificial neural network (ANN) [6] and support vector machine (SVM) will be adopted [7–9].

In this paper, we use multiplicative ARIMA, ARIMA with seasonal part, to model KPI time series with homogeneous nonstationarity.

Since ARIMA is a model evolving from ARMA, which can be regarded as an integration of autoregressive (AR) model and moving average (MA) model, AR, MA, and ARMA models are introduced first.

2.1 AR process

Autoregressive model has the form as follows:

$$x_t=\sum_{i=1}^{p}\phi_i x_{t-i}+\varepsilon_t, \tag{1}$$

where $\phi_p\neq 0$, and ε_t is a white noise process satisfying $\mathrm{WN}(0,\sigma^2)$.

Accordingly, the process defined by Eq. (1) is called an autoregressive process of order p, or more succinctly, an AR(p) process.

2.2 MA process

Moving average model has the form as follows:

$$Y_t=\varepsilon_t-\sum_{i=1}^{q}\theta_i\varepsilon_{t-i}, \tag{2}$$

where $\theta_q\neq 0$, and ε_t is a white noise process satisfying $\mathrm{WN}(0,\sigma^2)$.

The process defined by Eq. (2) is called moving average process of order q, which is abbreviated to MA(q).

2.3 Mixed ARMA process

Sometimes, a series can be modeled as autoregressive process, while some others can be modeled as moving average process. However, there are time series that cannot be modeled as pure AR or MA due to the too many parameters required. In these cases, mixed ARMA model should be used where both autoregressive and moving average terms are in the model, which can be represented as

$$x_t=\sum_{i=1}^{p}\phi_i x_{t-i}-\sum_{j=0}^{q}\theta_j\varepsilon_{t-j}, \tag{3}$$

or

$$\phi(B)x_t=\theta(B)\varepsilon_t. \tag{4}$$

In Eq. (4), B stands for backward shift operator, which is defined by $Bx_t=x_{t-1}$.

A process defined by Eq. (3) or (4) is called mixed autoregressive moving average process of order (p,q), which is abbreviated to ARMA(p,q).

2.4 ARIMA process

AR, MA, and ARMA models described above are all appropriate for modeling stationary process. However, in reality, many time series are nonstationary. For those nonstationary time series, nevertheless exhibiting homogeneity, we can stationalize them by differencing operation. Let $\nabla=1-B$ denote the differencing operator so that ARIMA model can be written as

$$\phi(B)\nabla^d x_t=\theta(B)\varepsilon_t, \tag{5}$$

where d is an integer representing number of differencing operation, $\phi(B)$ is a stationary autoregressive operator, and $\theta(B)$ is a moving average operator. Thus, we can get the conclusion that the model represented by Eq. (5) corresponds to the assumption that the dth difference of the series can be represented by a stationary ARMA process. The process satisfying Eq. (5) is called an ARIMA process.

Similarly, if the nonstationary time series exhibit periodical homogeneity with period s, they can be stationalized by seasonal differencing. The seasonal operator is of the form $\nabla_s=1-B^s$. If a time series is only of seasonal nonstationary, the model can be represented as

$$\Phi_P(B^s)\nabla_s^D x_t=\Theta_Q(B^s)\omega_t, \tag{6}$$

where

$$\Phi_P(B^s)=1-\Phi_1B^s-\Phi_2B^{2s}-\Phi_3B^{3s}-\cdots-\Phi_PB^{Ps},$$

and

$$\Theta_Q(B^s)=1-\Theta_1B^s-\Theta_2B^{2s}-\Theta_3B^{3s}-\cdots-\Theta_QB^{Qs}.$$

If time series have both trendy and seasonal part, integrating Eqs. (5) and (6), we can get the following

model:

$$\phi(B)\Phi_P(B^s)\nabla^d\nabla_s^D Y_t = \theta(B)\Theta_Q(B^s)\varepsilon_t. \quad (7)$$

The resulting multiplicative ARIMA process will be of order $(p,d,q)\times(P,D,Q)_s$.

3 Time series prediction based on ARIMA with associated confidence interval

For some KPIs such as traffic volume or call attempts, the baseline can be obtained according to time series prediction using ARIMA. The threshold then can be achieved based on the confidence interval under a dedicated confidence degree $1-\alpha$ by using the hypothesis that training residual satisfying white noise follows normal distribution.

Take the hourly traffic of mobile service switching center (MSC) for example. The traffics in different hours in a day are usually different. Because PPM takes hour as a unit, we predict hourly traffic using traffics at the same time in the past days. For example, if we want to predict the traffic at 9:00 tomorrow, then the history values are traffic data at 9:00 of today and yesterday and so on. Here, we use traffic values at 9:00 from 2007-7-4 to 2007-8-17, which is totally 45 consecutive items, as training data.

There are three steps to make the time series prediction using ARIMA. First, the ARIMA model should be identified. Second, the parameters should be estimated for the identified class of ARIMA models. Third, the prediction for future values should be made by using history values. In this paper, the associated confidence interval of prediction value should also be computed under given confidence degree.

The three steps above will be discussed in detail as follows.

3.1 Model identification

To get the ARIMA model fit for a time series, the order of the model $(p,d,q)\times(P,D,Q)_s$ should be decided at the first stage. Thus, an appropriate subclass of models can be identified from the general seasonal multiplicative ARIMA family denoted by Eq. (7). This stage is usually called model identification. By calculating the autocorrelation function, we can justify if the series exhibit trend or if the series shows seasonal pattern with period s. Then, we can make the series stationary by using a differencing operator or seasonal differencing operator. For stationary time series, there are several model identification criteria including AIC, BIC, F-test, etc.

In an actual application, especially when the series length of available data is not very large, the orders P, Q, p, and q would rarely need to be greater than 1. Therefore, at this stage of model identification, we should select the best model from all the combinations of $(p,d,q)\times(P,D,Q)_s$ where all the orders are between zero and one according to AIC.

3.2 Parameter estimation

After the appropriate model $(p,d,q)\times(P,D,Q)_s$ has been identified, the parameters would be estimated. There are several estimation algorithms, such as Yule-Walker equation, minimum mean squared error (MMSE), and maximum likelihood estimation (MLE). MLE is used in this paper.

3.3 Prediction with associated confidence interval

After the model is built up, it can be used to forecast future values based on history data. There are two types of forecasting, one-step ahead forecasting, and multisteps ahead forecasting. To guarantee the precision of forecasting, one-step ahead forecasting is adopted.

Actual value at time τ_0+ih is denoted by x_i, and prediction value for the same time is denoted by $\hat{x}_i$, the equation

$$x_i = \hat{x}_i + \varepsilon_i$$

holds, where ε_i is the prediction error and is white noise process following normal distribution, $\varepsilon_i \sim N(0,\sigma^2)$. Therefore, the actual value at the monitoring time point x_{n+1} and prediction value $\hat{x}_{n+1}$ satisfy $x_{n+1} \sim N(\hat{x}_{n+1},\sigma^2)$, which is equivalent to $\frac{x_{n+1}-\hat{x}_{n+1}}{\sigma} \sim N(0,1)$. Using a normal distribution table, we can compute the confidence interval of actual value under a given confidence degree $1-\alpha$, and the following equation is established:

$$P\left\{-z_{\alpha/2} \leqslant \frac{x_{n+1}-\hat{x}_{n+1}}{\sigma} \leqslant z_{\alpha/2}\right\} = 1-\alpha.$$

If the number of training items is more than 50, the training sample standard deviation

$$S = \sqrt{\frac{1}{k-1}\sum_{i=1}^{k}\left(\varepsilon_i - \overline{\varepsilon_i}\right)^2}$$

can be regarded as equal to the population standard deviation σ. Let a confidence degree be $1-\alpha = 0.95$, so that we can get the confidence interval of x_{n+1} as

$$(\hat{x}_{n+1}-2\sigma, \hat{x}_{n+1}+2\sigma) = (\hat{x}_{n+1}-2S, \hat{x}_{n+1}+2S).$$

If the confidence degree is $1-\alpha = 0.97$, the confidence interval will be $(\hat{x}_{n+1}-3\sigma, \hat{x}_{n+1}+3\sigma)$, which can be substituted with $(\hat{x}_{n+1}-3S, \hat{x}_{n+1}+3S)$.

However, if the number of training items is less than 50, the value of both σ and S may vary significantly so that they can not be regarded as being equivalent. In this case,

we propose the following algorithm to get the confidence interval.

Since training residual $\{\varepsilon_i\}$ satisfies normal distribution, according to the linear characteristics of operation of random variables obeying normal distribution, its mean value denoted by $\bar{\varepsilon}$ also satisfies normal distribution:

$$\bar{\varepsilon} \sim N\left(0, \frac{\sigma^2}{n}\right).$$

The random variable $\varepsilon_{n+1} - \bar{\varepsilon}$ will also satisfies the following normal distribution:

$$\varepsilon_{n+1} - \bar{\varepsilon} \sim N\left(0, \left[1 + \frac{1}{n^2}\right]\sigma^2\right), \tag{8}$$

and Eq. (8) can be transformed to

$$U = \frac{\varepsilon_{n+1} - \bar{\varepsilon}}{\sqrt{\frac{n^2+1}{n^2}}\sigma} \sim N(0,1). \tag{9}$$

Considering the fact that the standard deviation S of training residual satisfying χ^2 distribution, a random variable V can be derived as follows:

$$V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1). \tag{10}$$

Integrating Eqs. (9) and (10), another random variable named Z satisfying t distribution can be obtained by

$$Z = \frac{U}{\sqrt{\frac{V}{n-1}}} = \sqrt{\frac{n^2}{n^2+1}}\frac{(\varepsilon_{n+1} - \bar{\varepsilon})}{S} \sim t(n-1). \tag{11}$$

In Eq. (11), only ε_{n+1} is unknown, it can be transformed to

$$P\left\{\bar{\varepsilon} - \sqrt{\frac{n^2+1}{n^2}}t_{\alpha/2}(n-1)S \leqslant \varepsilon_{n+1} \leqslant \bar{\varepsilon} + \sqrt{\frac{n^2+1}{n^2}}t_{\alpha/2}(n-1)S\right\} = 1-\alpha. \tag{12}$$

Therefore, the confidence interval of x_{n+1} under confidence degree $1-\alpha$ will be

$$\left[\bar{\varepsilon} - \sqrt{\frac{n^2+1}{n^2}}t_{\alpha/2}(n-1)S + \hat{x}_{n+1}, \bar{\varepsilon} + \sqrt{\frac{n^2+1}{n^2}}t_{\alpha/2}(n-1)S + \hat{x}_{n+1}\right].$$

Thus, the former value is taken as the lower threshold, and the latter is the upper threshold with the confidence degree $1-\alpha$.

4 Anomaly detection for traffic series using ARIMA

Now, we have an hourly traffic load series at 9:00 from 2007-7-4 to 2007-8-17 with length of 45, as shown in Table 1. From the first item, we always take 37 consecutive items as training data to predict the next item with the associated confidence interval.

Table 1 Hourly traffic load series at 9:00 from 2007-7-4 to 2007-8-17

serial number	traffic load	serial number	traffic load
1	2390.78	24	2255.08
2	2339.33	25	2059.54
3	2338.28	26	2126.38
4	2180.47	27	2224.18
5	2123.48	28	2195.08
6	2474.97	29	2291.37
7	2495.80	30	2215.82
8	2416.97	31	2255.40
9	2363.72	32	2072.70
10	2373.28	33	2026.81
11	2140.83	34	2232.00
12	2109.89	35	2260.35
13	2337.10	36	2228.08
14	2254.07	37	2209.13
15	2224.28	38	2238.39
16	2228.55	39	2104.45
17	2252.30	40	2098.25
18	2121.81	41	2295.53
19	2075.98	42	2239.32
20	2331.03	43	2273.12
21	2278.63	44	2266.90
22	2255.06	45	2101.47
23	2264.72		

We stationalize the original traffic time series by differencing operation. The autocorrelation function of the original traffic series is depicted in Fig. 1. This figure reveals that there is no trend part but a marked seasonal pattern with periodicity of 7, so a seasonal difference should be carried out and denotes the resulting series by $\{y_i\}$, where $y_i = \nabla_7 x_i = (1-B^7)x_i$. Then, the stationary of $\{y_i\}$ is explored by calculating autocorrelation function. The result is shown in Fig. 2, where the seasonally differenced series $\{y_i\}$ is stationary.

We model the training data using ARIMA $(p,0,q) \times (P,1,Q)_7$ model and identify the model using AIC. Then,

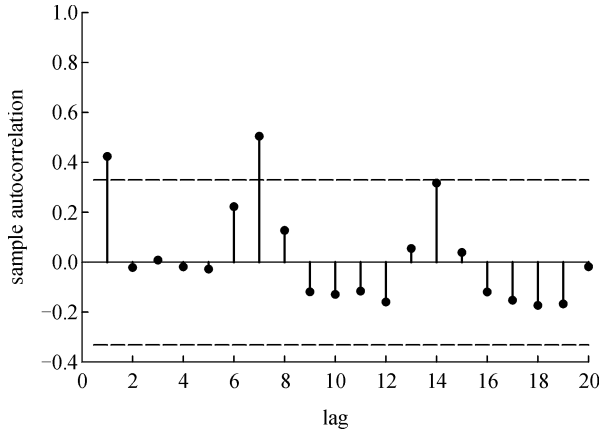

Fig. 1 Autocorrelation function of $\{x_i\}$

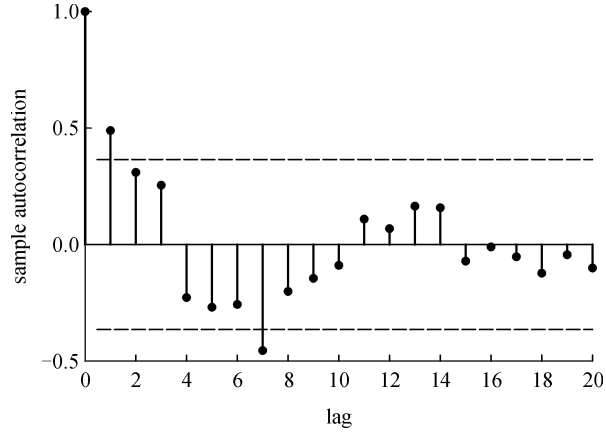

Fig. 2 Autocorrelation function of $\{y_i\}$

all the combinations of $(p,0,q)\times(P,1,Q)_7$ where each order is 0 or 1 were tested and ensured that they are not zeros at the same time and select the combination that makes AIC the least. The model $(1,0,1)\times(1,1,0)_7$ is selected, which means that for the original time series $\{x_i\}$ $(1,0,1)\times(1,1,0)_7$ is the best model.

We make a one-step ahead prediction as the baseline and calculate the associated confidence interval as the threshold. The one-step ahead prediction value is 2249.6, which was taken as the hourly traffic for 9:00, 2007-8-10. The sample standard deviation is $S=42.03$. With the confidence degree $1-\alpha=0.95$, we have

$$t_{\alpha/2}(n-1)=2.05,$$

and

$$\sqrt{\frac{n^2+1}{n^2}}t_{\alpha/2}(n-1)S=86.212.$$

With the confidence degree $1-\alpha=0.95$, the confidence interval is

$$[2249.6+2.77-86.212,\ 2249.6+2.77+86.212]$$

$$=[2166.158,\ 2338.582].$$

Therefore, we reach conclusion that the actual traffic value for 2007-8-10 9:00 is 2238.39, which is among the confidence interval [2166.158, 2338.582]. This means that during this period, the network is in normal state.

Using the method above, the prediction with associated confidence interval can be made from 2007-8-10 until 2007-8-17. The result is shown in Table 2 and plotted in Fig. 3.

We come to the following conclusions based on Table 2 and Fig. 3:

1) The actual traffics at 9:00 from 2007-8-10 till 2007-8-16 are all within the tolerance limit and all the values of absolute percent errors (APEs) are smaller than 3%. Figure 3 shows that in normal state, the proposed prediction approach is of high precision.

2) However, at 9:00 on 2007-8-17, the actual traffic, 2101.47, falls outside tolerance limit [2156.72, 2364.56]. In this status, a performance alarm should be launched. This abnormal status has been confirmed by the fact that during this period, there was something wrong with IC2 module in BSU unit, which caused some call setup failure. That is why the traffic was so low then. 2007-8-17 is Friday when traffic is usually the heaviest in the week. However, Fig. 3 shows us an opposite case that the traffic on 2007-8-17 is the lightest in the week, which is obviously abnormal.

Table 2 Prediction for 9:00 2007-8-10 till 9:00 2007-8-17

	2007-8-10	2007-8-11	2007-8-12	2007-8-13	2007-8-14	2007-8-15	2007-8-16	2007-8-17
actual	2238.39	2104.45	2098.25	2295.53	2239.32	2273.12	2266.9	2101.47
prediction	2249.578	2056.227	2069.271	2233.799	2248.385	2255.076	2240.787	2245.69
absolute error (AE)	11.188	48.223	28.979	61.731	9.065	18.044	26.113	144.22
APE	0.500%	2.291%	1.381%	2.689%	0.405%	0.794%	1.152%	6.863%
lower bound	2166.16	1974.04	1984.45	2119.73	2140.27	2151.33	2138.71	2156.72
upper bound	2338.58	2138.41	2154.09	2347.86	2356.49	2358.82	2342.85	2364.56

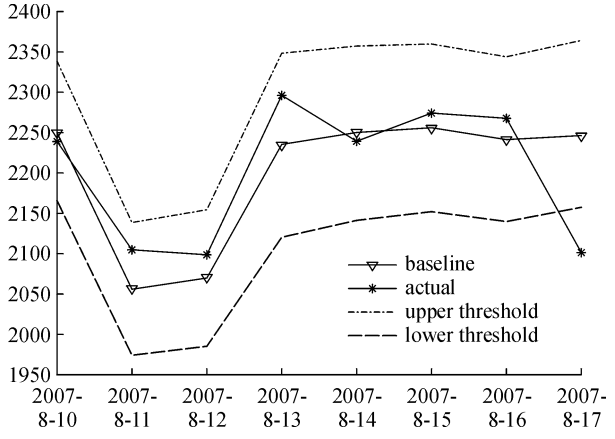

Fig. 3 Prediction value, actual value, thresholds for traffic from 2007-8-10 to 2007-8-17

5 Conclusion

The prediction of baseline and threshold for KPIs is a crucial issue for proactive performance monitoring of telecommunication network. In this paper, the approach of time series prediction with the associated confidence interval using ARIMA was adopted to solve this problem. Furthermore, under the assumption that a white noise process follows normal distribution, the associated confidence interval of prediction can be computed under any given confidence degree $1-\alpha$ by constructing random variable satisfying t distribution. Experimental results show that using the proposed modeling method, a quite precise prediction and fluctuation range can be achieved.

Acknowledgements This work was supported by the National Key Technologies R&D program of China during the 11th Five-Year Plan Period (No. 2006BAH02A03).

References

1. Hellerstein J L, Zhang F, Shahabuddin P. An approach to predictive detection for service management. In: Proceedings of the 6th IFIP/IEEE International Symposium on Integrated Network Management, 1999, Sloman M, Mazumdar S, Lupu E, Eds. New York: IEEE Publishing, 1999, 309–322
2. Feather F, Siewiorek D, Maxion R. Fault detection in an Ethernet network using anomaly signature matching. ACM SIGCOMM Computer Communication Review, 1993, 23(4): 279–288
3. Ho L L, Cavuto D J, Papavassiliou S, Zawadki A G. Adaptive and automated detection of service anomalies in transaction-oriented WANs: network analysis, algorithms, implementation, and deployment. IEEE Journal on Selected Areas in Communications, 2000, 18 (5): 744–757
4. Li J, Liu X X, Han Z J. Research on the ARMA-based traffic prediction algorithm for wireless sensor network. Journal of Electronics and Information Technology, 2007, 29(5): 1224–1227 (in Chinese)
5. Cadzow J A. ARMA time series modeling: an effective method. IEEE Transactions on Aerospace and Electronic Systems, 1983, AES-19 (1): 49–58
6. Versace M, Bhatt R, Hinds O, Shiffer M. Predicting the exchange traded fund DIA with a combination of genetic algorithms and neural networks. Expert Systems with Applications, 2004, 27(3): 417–425
7. Mukherjee S, Osuna E, Girosi F. Nonlinear prediction of chaotic time series using support vector machines. In: Proceedings of the 1997 IEEE Workshop on Neural Networks for Signal Processing, 1997, 511–520
8. Shi Z W, Han M. Support vector echo-state machine for chaotic time-series prediction. IEEE Transactions on Neural Networks, 2007, 18 (2): 359–372
9. Cao L J, Tay F E H. Support vector machine with adaptive parameters in financial time series forecasting. IEEE Transactions on Neural Networks, 2003, 14(6): 1506–1518